

# Quantitative compactness estimates for Hamilton-Jacobi equations

Fabio Ancona\*, Piermarco Cannarsa†, Khai T. Nguyen‡

March 19, 2014

## Abstract

We study quantitative compactness estimates in  $\mathbf{W}_{loc}^{1,1}$  for the map  $S_t$ ,  $t > 0$  that associates to every given initial data  $u_0 \in \text{Lip}(\mathbb{R}^N)$  the corresponding solution  $S_t u_0$  of a Hamilton-Jacobi equation

$$u_t + H(\nabla_x u) = 0, \quad t \geq 0, \quad x \in \mathbb{R}^N,$$

with a uniformly convex Hamiltonian  $H = H(p)$ . We provide upper and lower estimates of order  $1/\varepsilon^N$  on the the Kolmogorov  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  of the image through the map  $S_t$  of sets of bounded, compactly supported initial data. Estimates of this type are inspired by a question posed by P.D. Lax [18] within the context of conservation laws, and could provide a measure of the order of “resolution” of a numerical method implemented for this equation.

**Key words:** Hamilton-Jacobi equations, Hopf-Lax semigroup, compactness estimates, Kolmogorov entropy, semiconcave functions

**MSC Subject classifications:** 49L05, 47H20, 49L20, 47H08

## 1 Introduction

Consider a first-order Hamilton-Jacobi equation

$$u_t(t, x) + H(\nabla_x u(t, x)) = 0, \quad t \geq 0, \quad x \in \mathbb{R}^N, \quad (1)$$

where  $u = u(t, x)$ ,  $\nabla_x u = (u_{x_1}, \dots, u_{x_N})$ , and  $H : \mathbb{R}^N \rightarrow \mathbb{R}$  is a smooth Hamiltonian. It is well-known that, because of the nonlinear dependence of the characteristic speeds on the gradient of the solution, in general classical solutions  $u(t, x)$  of the Cauchy problem for (1) develop singularities of  $\nabla_x u(t, x)$  in finite time, no matter how smooth the initial data

$$u(0, \cdot) = u_0 \quad (2)$$

are assumed to be. To cope with this difficulty, M.G. Crandall and P.-L. Lions introduced in [10] the notion of viscosity solution, a generalized solution of (1), which allows to establish global existence, uniqueness and stability results for the Cauchy problem (1)-(2), under suitable assumption on  $H$ . We refer to [7] for a review of the concept of viscosity solution and the related theory for equation of type (1) that has been developed in the last thirty years.

---

\*Dipartimento di Matematica, Università di Padova, Via Trieste 63, 35121 Padova, Italy

†Dipartimento di Matematica, Università di Roma 'Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy

‡Department of Mathematics, Penn State University, University Park, Pa. 16802, U.S.A.

The Hamiltonian  $H$  is required here to satisfy the **Standing Assumption**:

**(H1)**  $H \in C^2(\mathbb{R}^N)$  and is uniformly convex, i.e.

$$D^2H(p) \geq \alpha \cdot \mathbb{I}_N \quad \forall p \in \mathbb{R}^N,$$

where  $\alpha$  is a positive constant,  $\mathbb{I}_N$  is the  $N \times N$  identity matrix, and the inequality is understood in the sense that  $D^2H(p) - \alpha \cdot \mathbb{I}_N$  is a positive semidefinite matrix.

The assumption **(H1)** guarantees that, if the initial data  $u_0 : \mathbb{R}^N \rightarrow \mathbb{R}$  is Lipschitz continuous and bounded, the Cauchy problem (1)-(2) admits a unique viscosity solution  $u(t, x)$  which is Lipschitz continuous and semiconcave in  $x$  with semiconcavity constant  $1/(\alpha t)$ . This means that  $x \mapsto u(t, x) - 1/(2\alpha t)|x|^2$  is a concave function. In turn, this fact implies that  $u(t, \cdot)$  is almost everywhere twice differentiable and that  $\nabla_x u(t, \cdot)$  has locally bounded variation, i.e. that the distributional Hessian  $D_x^2 u(t, \cdot)$  is a symmetric matrix of Radon measures.

Furthermore, one can define a semigroup of viscosity solutions of (1)

$$\{S_t : \text{Lip}(\mathbb{R}^N) \rightarrow \text{Lip}(\mathbb{R}^N)\}_{t \geq 0}$$

that associates to every initial data  $u_0 \in \text{Lip}(\mathbb{R}^N)$  the unique viscosity solution  $S_t u_0 := u(t, \cdot)$  of the corresponding Cauchy problem (1)-(2). It is not difficult to see that the semigroup map  $S_t$  is continuous when it is restricted to subsets of  $\text{Lip}(\mathbb{R}^N)$  bounded in  $\mathbf{W}^{1,\infty}$ , taking the  $\mathbf{W}_{loc}^{1,1}$ -topology on  $\text{Lip}(\mathbb{R}^N)$  (cfr. Proposition 5 in Section 2). Moreover, thanks to the uniform semiconcavity constant of  $S_t u_0$ , for  $u_0 \in \text{Lip}(\mathbb{R}^N)$ , applying Helly's compactness theorem and a Poincaré inequality for BV-functions, one can show that the restriction of  $S_t$ ,  $t > 0$ , to such sets is compact with respect to the  $\mathbf{W}_{loc}^{1,1}$ -topology. This property reflects the irreversibility feature of the equation (1) when the Hamiltonian  $H$  satisfies the convexity assumption **(H1)**.

The aim of this paper is to provide a quantitative estimate of this regularizing effect of the semigroup map. Namely, having in mind a question posed by P.D. Lax [18] within the context of conservation laws, we wish to estimate the Kolmogorov  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  of the image through the map  $S_t$  of sets of bounded, compactly supported initial data  $\mathcal{C} \subset \text{Lip}(\mathbb{R}^N)$  of the form

$$\mathcal{C}_{[L,M]} := \left\{ u_0 \in \text{Lip}(\mathbb{R}^N) \mid \text{supp}(u_0) \subset [-L, L]^N, \text{Lip}[u_0] \leq M \right\}. \quad (3)$$

Actually, since the solution of the Cauchy problem for (1) with zero initial data is the function  $u(t, x) = -t \cdot H(0)$ , it will be convenient to analyze the Kolmogorov  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  of the translated set  $S_t(\mathcal{C}) + t \cdot H(0)$ , with  $\mathcal{C}$  as in (3). We recall the notion of  $\varepsilon$ -entropy introduced by A. Kolmogorov [16]:

**Definition 1.** Let  $(X, d)$  be a metric space and let  $K$  be a totally bounded subset of  $X$ . For  $\varepsilon > 0$ , let  $\mathcal{N}_\varepsilon(K|X)$  be the minimal number of sets in a cover of  $K$  by subsets of  $X$  having diameter no larger than  $2\varepsilon$ . Then the  $\varepsilon$ -entropy of  $K$  is defined as

$$\mathcal{H}_\varepsilon(K|X) := \log_2 \mathcal{N}_\varepsilon(K|X).$$

Throughout the paper, we will call  $\varepsilon$ -cover a cover of  $K$  by subsets of  $X$  having diameter no larger than  $2\varepsilon$ .

Entropy numbers play a central role in various areas of information theory and statistics as well as of ergodic and learning theory. In the present setting, this concept could provide a measure of the order of “resolution” and of the “complexity” of a numerical scheme, as suggested in [17]. Roughly speaking, the order of magnitude of the  $\varepsilon$ -entropy should indicate the minimum

number of operations that one should perform in order to obtain an approximate solution with a precision of order  $\varepsilon$  with respect to the considered topology.

In this paper we provide both upper and lower bounds of order  $1/\varepsilon^N$  on the  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  of  $S_t(\mathcal{C}) + t \cdot H(0)$ , for sets  $\mathcal{C}$  as in (3), thus showing that such an  $\varepsilon$ -entropy is of size  $\approx 1/\varepsilon^N$ . Without loss of generality, we will assume that the Hamiltonian satisfies further

$$(H2) \quad \nabla H(0) = 0,$$

otherwise the transformations  $x \rightarrow x + t\nabla H(0)$  and  $H(p) \rightarrow H(p) - \langle \nabla H(0), p \rangle$  reduce the general case to this one. Specifically, we prove the following

**Theorem 1.** *Let  $H : \mathbb{R}^N \rightarrow \mathbb{R}$  be a function satisfying the assumptions (H1)-(H2) and  $\{S_t\}_{t \geq 0}$  be the semigroup of viscosity solutions generated by (1) on the domain  $\text{Lip}(\mathbb{R}^N)$ . Then, given  $L, M, T > 0$ , for every  $\varepsilon > 0$  sufficiently small the following estimates hold:*

$$\mathcal{H}_\varepsilon\left(S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0) \mid \mathbf{W}^{1,1}(\mathbb{R}^N)\right) \leq \Gamma_{[L,M,N,T]}^+ \cdot \frac{1}{\varepsilon^N} \quad (4)$$

with

$$\Gamma_{[L,M,N,T]}^+ := \omega_N^N \cdot \left(4N \cdot \left(1 + M + (1/(\alpha T) + 1) \cdot l_{[L,M,T]}\right)\right)^{4N^2} \quad (5)$$

$$l_{[L,M,T]} := L + T \cdot \sup_{|p| \leq M} |\nabla H(p)|, \quad (6)$$

$\alpha$  being the constant appearing in (H1) and  $\omega_N$  denoting the Lebesgue measure of the unit ball of  $\mathbb{R}^N$ , and

$$\mathcal{H}_\varepsilon\left(S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0) \mid \mathbf{W}^{1,1}(\mathbb{R}^N)\right) \geq \Gamma_{[L,N,T]}^- \cdot \frac{1}{\varepsilon^N} \quad (7)$$

with

$$\Gamma_{[L,N,T]}^- := \frac{1}{8 \cdot \ln 2} \left( \frac{\omega_N}{192(N+1) \cdot \|D^2 H(0)\| \cdot T} \right)^N \cdot \left( \frac{L}{4} \right)^{N(N+1)}. \quad (8)$$

In the one dimensional case ( $N = 1$ ) the above estimates can be easily obtained recalling the well-known fact (e.g. see [15]) that  $u(t, x)$  is a viscosity solution of (1) if and only if its space derivative  $v(t, x) := u_x(t, x)$  is an entropy weak solution of the conservation law

$$v_t + H(v)_x = 0, \quad (9)$$

and relying on the same type of estimates established in [4, 12] for scalar conservation laws. In fact, denoting with  $\tilde{S}_t$  the semigroup map generated by (9), observe that any  $\varepsilon$ -cover in  $\mathbf{W}^{1,1}$  for a translated set  $S_t(\mathcal{C}) + t \cdot H(0)$  of solutions to (1) at time  $t$ , with initial data in  $\mathcal{C}$ , provides also an  $\varepsilon$ -cover in  $\mathbf{L}^1$  for the set  $\tilde{S}_t(\mathcal{C}')$  of solutions to (9) at time  $t$ , with initial data in  $\mathcal{C}' := \{u' \mid u \in \mathcal{C}\}$ . Thus, applying [4, Theorem 1.3] one derives the lower bound  $\mathcal{H}_\varepsilon(S_t(\mathcal{C}) + t \cdot H(0) \mid \mathbf{W}^{1,1}) \geq \mathcal{H}_\varepsilon(S_t(\mathcal{C}') \mid \mathbf{L}^1) \gtrsim \frac{L^2}{|H''(0)| \cdot t} \cdot \frac{1}{\varepsilon}$ , which is of the same size as the one provided by  $\Gamma_{[L,1,t]}^- \cdot \frac{1}{\varepsilon}$  in (7). On the other hand, invoking a Poincaré inequality, one can easily adapt the construction performed in [12] of an  $\varepsilon$ -cover in  $\mathbf{L}^1$  of  $\tilde{S}_t(\mathcal{C}')$  to produce an  $\varepsilon$ -cover in  $\mathbf{W}^{1,1}$  of  $S_t(\mathcal{C}) + t \cdot H(0)$  with the same number of elements. As a consequence, we derive an upper bound on  $\mathcal{H}_\varepsilon(S_t(\mathcal{C}) + t \cdot H(0) \mid \mathbf{W}^{1,1})$  of the same order as the one established in [12, Theorem 2.2] (cfr. also [4, Remark 1.4]) which, in turn, is of the same size as the one provided by  $\Gamma_{[L,M,1,t]}^+ \cdot \frac{1}{\varepsilon}$  in (4).

When the space dimension is greater than one we can no more rely on the equivalence between the theory of Hamilton-Jacobi equations and that of hyperbolic conservation laws. Indeed,

in this case, the gradient of a viscosity solution turns out to be (at least formally) a solution of a non-strictly hyperbolic system in several space variables, while the available compactness estimates for systems of conservation laws concern only the class of strictly hyperbolic systems in one space variable [5, 6]. Nevertheless, we shall implement some of the ideas originated in the works [4, 12] to prove Theorem 1. However, in order to handle the higher dimensional case, one needs new ideas which exploit specific properties of the viscosity solutions of (1) as well as the geometrical theory of monotone functions of several variables.

Towards the derivation of the upper bound stated in (i), we observe that for any given viscosity solution  $u(t, x)$ , letting  $D_x^+ u$  denote a generalized space gradient of  $u$  (cfr. Definition 3), the semiconcavity property of  $u$  ensures that the map  $x \mapsto D_x^+ u(t, x) - \frac{x}{\alpha t}$  is a monotone decreasing multifunction on  $\mathbb{R}^N$ . Next, relying on a Poincaré inequality, we provide an upper bound on the  $\varepsilon$ -entropy in  $\mathbf{L}^1$  for a class of monotone decreasing multifunctions with uniformly bounded total variation, defined on a bounded domain of  $\mathbb{R}^N$ . In turn, such a bound yields estimate (4) on the  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  of  $S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0)$ , again by Poincaré's inequality.

The lower bounds on  $H_\varepsilon(S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0))$  are obtained in two steps adopting a similar strategy as the one pursued in [4].

1. We consider a class  $\mathcal{SC}_{[K]}$  of semiconcave functions with semiconcavity constant  $K$ , defined on a bounded domain, and we establish a controllability type result for the elements of such a class, up to a translation by a fixed map. Namely, employing the Hopf-Lax formula for the viscosity solutions to (1) we prove that, at any given time  $T > 0$ , every element of  $\mathcal{SC}_{[K]} - T \cdot H(0)$  can be obtained as the value  $u(T, \cdot)$  of a classical solution of (1), with initial data in  $\mathcal{C}_{[L,M]}$ , provided that the semiconcavity constant  $K$  is sufficient small. Since a classical solution must coincide with the unique viscosity solution of the corresponding Cauchy problem, this proves that  $\mathcal{SC}_{[K]} - T \cdot H(0) \subset S_T(\mathcal{C}_{[L,M]})$ .
2. We introduce a one-parameter class of semiconcave functions  $\mathcal{U}_n \subset \mathcal{SC}_{[K]}$  defined as combinations of suitable bump functions and, by a combinatorial argument, we provide an optimal estimate (w.r.t. parameter  $n$ ) of the maximum number of functions in  $\mathcal{U}_n$  at distance  $\leq \varepsilon$  w.r.t. the  $\mathbf{W}^{1,1}$ -metric. This estimate yields a lower bound on the  $\varepsilon$ -entropy of  $\mathcal{U}_n$ , from which we recover (7) relying on the result of point 1.

The paper is organized as follows. In Section 2, we collect preliminary results and definitions concerning semiconcave functions and Hamilton-Jacobi equations. In Section 3, after deriving further properties of the viscosity solutions of Hamilton-Jacobi equations, we provide an upper bound on the  $\varepsilon$ -entropy in  $\mathbf{L}^1$  for a class of monotone multifunctions. Relying on this result, we next establish an upper bound on the  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  for a class of semiconcave functions, which yields the upper bound stated in Theorem 1-(i). In Section 4, we carry out the analysis described in the above two steps to obtain the lower bound stated in Theorem 1-(ii).

## 2 Notation and preliminaries

Let  $N \geq 1$  be an integer. Throughout the paper we shall denote by:

- $|\cdot|$  the Euclidean norm in  $\mathbb{R}^N$ ,
- $\langle \cdot, \cdot \rangle$  the Euclidean inner product in  $\mathbb{R}^N$ ,
- $[x, y]$  the segment joining two points  $x, y \in \mathbb{R}^N$ ,
- $B(x_0, r)$  the open ball of  $\mathbb{R}^N$  with radius  $r > 0$  and centered at  $x_0$ ,

- $\#(S)$  the number of elements of any finite set  $S$ ,
- $\text{Vol}(D)$  the Lebesgue measure of a measurable set  $D \subset \mathbb{R}^N$ ,
- $\omega_N := \text{Vol}(B(0, 1)) = \frac{\pi^{N/2}}{\Gamma(N/2+1)}$  the Lebesgue measure of the unit ball of  $\mathbb{R}^N$ ,
- $\|A\|$  the usual operator norm of the  $N \times N$  matrix  $A$ ,
- $\text{Lip}(\mathbb{R}^N)$  the space of all Lipschitz continuous functions  $f : \mathbb{R}^N \rightarrow \mathbb{R}$ , and by  $\text{Lip}[f]$  the Lipschitz seminorm of  $f$ ,
- $\text{supp}(u)$  the support of  $u \in \text{Lip}(\mathbb{R}^N)$ , that is, the closure of  $\{x \in \mathbb{R}^N \mid u(x) \neq 0\}$ ,
- $\mathbf{L}^1(D)$ , with  $D \subset \mathbb{R}^N$  a measurable set, the Lebesgue space of all (equivalence classes of) summable functions on  $D$ , equipped with the usual norm  $\|\cdot\|_{\mathbf{L}^1(D)}$ ,
- $\mathbf{L}^\infty(D)$ , with  $D \subset \mathbb{R}^N$  a measurable set, the space of all essentially bounded functions on  $D$ , and by  $\|u\|_{\mathbf{L}^\infty(D)}$  the essential supremum of a function  $u \in \mathbf{L}^\infty(D)$  (we shall use the same symbol in case  $u$  is vector-valued),
- $\mathbf{W}^{1,1}(\Omega)$ , with  $\Omega$  a convex domain in  $\mathbb{R}^N$ , the Sobolev space of functions with summable first order distributional derivatives, and by  $\|\cdot\|_{\mathbf{W}^{1,1}(\Omega)}$  its norm,
- $\mathbf{W}_0^{1,1}(\Omega)$ , with  $\Omega$  a convex domain in  $\mathbb{R}^N$ , the Sobolev space of functions  $F \in \mathbf{W}^{1,1}(\Omega)$  with zero trace on the boundary  $\partial\Omega$ ,
- $BV(\Omega, \mathbb{R}^N)$ , with  $\Omega$  a domain in  $\mathbb{R}^N$ , the space of all vector-valued functions  $F : \Omega \rightarrow \mathbb{R}^N$  of bounded variation (that is, all  $F \in \mathbf{L}^1(\Omega, \mathbb{R}^N)$  such that the first partial derivatives of  $F$  in the sense of distributions are measures with finite total variation in  $\Omega$ ).

Moreover  $[a] := \max\{z \in \mathbb{Z} \mid z \leq a\}$  denotes the integer part of  $a$ .

## 2.1 Semiconcave and monotone functions in $\mathbb{R}^N$ and Poincaré inequalities

We collect here some basic definitions and properties of semiconcave and monotone functions in  $\mathbb{R}^N$  that will be used in the paper. We refer to [9] and [2] for a general introduction to the respective theories.

**Definition 2.** A continuous function  $u : \Omega \rightarrow \mathbb{R}$ , with  $\Omega \subset \mathbb{R}^N$ , is called *semiconcave* if there exists  $K > 0$  such that

$$u(x+h) + u(x-h) - 2u(x) \leq K|h|^2, \quad (10)$$

for all  $x, h \in \mathbb{R}^N$  such that  $[x-h, x+h] \subset \Omega$ . When this property holds true, we also say that  $u$  is *semiconcave* in  $\Omega$  with constant  $K$ , and call  $K$  a *semiconcavity constant* for  $u$ .

- We say that  $u$  is *semiconvex* (with constant  $-K$ ) if  $-u$  is *semiconcave* (with constant  $K$ ).
- We say that  $u : \Omega \rightarrow \mathbb{R}$ , with  $\Omega \subset \mathbb{R}^N$  open, is *locally semiconcave* (or *locally semiconvex*) if  $u$  is *semiconcave* (semiconvex) in every compact set  $A \subset\subset \Omega$ .

**Remark 1.** The notion of semiconcavity introduced here is the most commonly used in the literature. A more general definition of semiconcavity can be found in [9]. It is easy to see that a function  $u$  is semiconcave in  $\Omega$  with constant  $K$  if and only if the function

$$\tilde{u}(x) = u(x) - \frac{K}{2}|x|^2 \quad (x \in \Omega)$$

is concave. Moreover, any continuously differentiable map  $u : \Omega \rightarrow \mathbb{R}$  that has a Lipschitz continuous gradient  $\nabla u$  with Lipschitz constant  $K$  is semiconcave with constant  $2K$ .

Semiconcave functions share some well-known properties of concave functions (see [9, Theorem 2.1.7, Theorem 2.3.1] and [2, Proposition 5711]) stated in the following

**Theorem 2.** *Let  $\Omega \subseteq \mathbb{R}^N$  open and  $u : \Omega \rightarrow \mathbb{R}$  be locally semiconcave. Then, the following properties hold true:*

- (i)  *$u$  is locally Lipschitz continuous.*
- (ii) *(Alexandroff's Theorem)  $u$  is almost everywhere twice differentiable.*
- (iii) *The gradient of  $u$ , defined almost everywhere in  $\Omega$ , belongs to  $BV_{loc}(\Omega, \mathbb{R}^N)$ . Moreover, if  $u$  is semiconcave in  $\Omega$  with constant  $K$ , then*

$$D^2u \leq K \cdot \mathbb{I}_N \mathcal{L}^N \quad (11)$$

*in the sense of symmetric matrix-valued measures.*

We shall adopt the notation  $Du$  for the distributional gradient of a semiconcave function  $u$ . A notion of generalized gradient that is specially fit to viscosity solutions is recalled in the following

**Definition 3.** *Let  $u : \Omega \rightarrow \mathbb{R}$ , with  $\Omega \subseteq \mathbb{R}^N$  open. For every  $x \in \Omega$ , the sets*

$$\begin{aligned} D^+u(x) &:= \left\{ p \in \mathbb{R}^N \mid \limsup_{y \rightarrow x} \frac{u(y) - u(x) - \langle p, y - x \rangle}{|y - x|} \leq 0 \right\}, \\ D^-u(x) &:= \left\{ p \in \mathbb{R}^N \mid \liminf_{y \rightarrow x} \frac{u(y) - u(x) - \langle p, y - x \rangle}{|y - x|} \geq 0 \right\}, \end{aligned} \quad (12)$$

*are called, respectively, the superdifferential and the subdifferential of  $u$  at  $x$ . Moreover,*

$$D^*u(x) := \left\{ p = \lim_{k \rightarrow \infty} \nabla u(x_k) \mid f \text{ is differentiable at } x_k \text{ and } x_k \rightarrow x \right\}, \quad (13)$$

*is called the set of reachable gradients of  $u$  at  $x$ .*

From definition (12) it follows that there holds

$$D^-u(x) = -D^+(-u)(x) \quad \forall x \in \Omega. \quad (14)$$

The superdifferential of a semiconcave function enjoys the properties stated in the following (see [9, Proposition 3.3.4, Theorem 3.3.6])

**Theorem 3.** *Let  $\Omega \subseteq \mathbb{R}^N$  open and  $u : \Omega \rightarrow \mathbb{R}$  be locally semiconcave. Then, the following properties hold true.*

- (i) *The superdifferential  $D^+u(x)$  is a compact, convex, nonempty set for all  $x \in \Omega$ .*
- (ii)  *$D^+u$  is an upper semicontinuous set-valued map, that is, if  $\{x_k\}$  is a sequence in  $\Omega$  converging to  $x$ , and if  $p_k \in D^+u(x_k)$  converges to a vector  $p \in \mathbb{R}^N$ , then  $p \in D^+u(x)$ .*
- (iii)  *$D^+u(x) = \text{co } D^*u(x)$  for all  $x \in \Omega$ , where  $\text{co}$  stands for the convex hull.*
- (iv)  *$D^+u(x)$  is a singleton if and only if  $u$  is differentiable at  $x$ .*

(v) If  $D^+u(x)$  is a singleton for every  $x \in \Omega$ , then  $u \in C^1(\Omega, \mathbb{R})$ .

**Remark 2.** Relying on the properties of the generalized gradients one can show that if a function  $u : \Omega \rightarrow \mathbb{R}$  ( $\Omega \subseteq \mathbb{R}^N$  open and convex) is both semiconcave and semiconvex in  $\Omega$  then  $u \in C^{1,1}(\Omega, \mathbb{R})$  (see [9, Corollary 3.3.8]).

In dealing with the map  $x \mapsto D^+u(x)$  it will be useful to recall the following notions for set-valued maps.

**Definition 4.** Let  $F : \mathbb{R}^N \rightarrow 2^{\mathbb{R}^N}$  be a multifunction, that is a map that associates with every point  $x \in \mathbb{R}^N$  some set  $F(x) \subset \mathbb{R}^N$ . We say that  $F$  is monotone decreasing if

$$\langle v_2 - v_1, x_2 - x_1 \rangle \leq 0, \quad \forall x_i \in \mathbb{R}^N, v_i \in F(x_i), i = 1, 2. \quad (15)$$

The set

$$\text{dom}(F) := \{x \in \mathbb{R}^N \mid F(x) \neq \emptyset\}$$

is called the domain of  $F$ . We say that  $F$  is univalued on some set  $A$  if  $F(x)$  consists of at most one point for every  $x \in A$ .

As observed in [2] (see Corollary 1.3(3) and Remark 2.3), any monotone decreasing multifunction  $F$  is bounded and almost everywhere univalued in every open set  $\Omega \subset \mathbb{R}^N$ , which is relatively compact in the interior of  $\text{dom}(F)$ . Therefore, we may regard the restriction of  $F$  to any such open set  $\Omega$  as an element of  $\mathbf{L}^\infty(\Omega, \mathbb{R}^N)$ . Actually, in [2, Proposition 5.1],  $F$  is shown to be a function of bounded variation on  $\Omega$  and the following upper bound on the total variation of its distributional derivative is provided.

**Proposition 1.** Let  $F : \mathbb{R}^N \rightarrow 2^{\mathbb{R}^N}$  be a monotone decreasing multifunction and  $\Omega \subset \mathbb{R}^N$  be an open set, relatively compact in the interior of  $\text{dom}(F)$ . Then, the restriction of  $F$  to  $\Omega$  (viewed as an element of  $\mathbf{L}^\infty(\Omega, \mathbb{R}^N)$ ) belongs to  $BV(\Omega, \mathbb{R}^N)$ . Moreover, setting  $F(\Omega) := \cup_{x \in \Omega} F(x)$ , there holds

$$|DF|(\Omega) \leq 2^{\frac{N}{2}} N^2 \omega_N [\text{diam}(\Omega) + \text{diam}(F(\Omega))]^N \quad (16)$$

where  $|DF|$  is the total variation of the (matrix-valued) Radon measure  $DF$ , and

$$\text{diam}(A) := \sup \{|x_2 - x_1| \mid x_i \in A\} \quad (A \subset \mathbb{R}^N).$$

We next recall further properties of semiconcave functions and of their superdifferentials (see [9, Theorem 2.1.7, Theorem 2.3.1, Proposition 3.3.10], [2, Corollary 1.4]).

**Proposition 2.** *Let  $\Omega \subseteq \mathbb{R}^N$  be open convex and  $u : \Omega \rightarrow \mathbb{R}$  be semiconcave with constant  $K$ . Then, the following properties hold.*

(i) *For every  $x, y \in \Omega$ , and for any  $p_x \in D^+u(x)$ ,  $p_y \in D^+u(y)$ , there holds*

$$\langle p_y - p_x, y - x \rangle \leq K |y - x|^2.$$

(ii) *The map  $x \mapsto D^+u(x) - Kx$  is a monotone decreasing multifunction.*

We conclude this paragraph recalling two Poincaré-type inequalities that will be used in the paper. The first one is valid for trace-zero  $\mathbf{W}^{1,1}$  functions (e.g. see [13, Theorem 3 in Section 5.6]), while the second one, based on [1, Theorem 3.2] and on [3, Proposition 3.2.1, Theorem 3.44], is satisfied by BV functions on convex domain.

**Theorem 4.** *(Poincaré inequalities) Let  $\Omega \subseteq \mathbb{R}^N$  be an open, bounded and convex set with Lipschitz boundary.*

(i) *If  $u \in \mathbf{W}_0^{1,1}(\Omega)$ , then*

$$\int_{\Omega} |u| dx \leq (\text{Vol}(\Omega))^{\frac{1}{N}} \int_{\Omega} |\nabla u| dx. \quad (17)$$

(ii) *If  $u \in BV(\Omega, \mathbb{R}^N)$ , then, letting*

$$u_{\Omega} := \frac{1}{\text{Vol}(\Omega)} \int_{\Omega} u(x) dx,$$

*denote the mean value of  $u$  over  $\Omega$ , there holds*

$$\int_{\Omega} |u - u_{\Omega}| dx \leq \frac{\text{diam}(\Omega)}{2} \cdot |Du|(\Omega), \quad (18)$$

*where  $|Du|$  is the total variation of the Radon measure  $Du$ .*

## 2.2 Hamilton-Jacobi equation

Consider the Hamilton-Jacobi equation (1) under the assumptions **(H1)**-**(H2)**. Observe that the lower bound on the Hessian matrix  $D^2H$  given in **(H1)** in particular implies the condition:

**(H1)'**  $H \in C^2(\mathbb{R}^N)$  and is a uniformly convex and coercive map, i.e.,

$$\lim_{|p| \rightarrow \infty} \frac{H(p)}{|p|} = +\infty.$$

Moreover, relying on **(H1)**, we have that

$$\exists m_0 > 0 \quad \text{s.t.} \quad \sup_{|p| \leq m_0} \|D^2H(p)\| \leq 2 \|D^2H(0)\|. \quad (19)$$

As we mentioned in the introduction, classical smooth solutions of (1) in general break down and Lipschitz continuous functions that satisfy (1) almost everywhere together with an initial condition (2) are not unique. To handle this problem, the following concept of solution was introduced in [10] (see also [11]) so to guarantee global existence and uniqueness results.



**Definition 5.** (*Viscosity solution*) We say that a continuous function  $u : [0, T] \times \mathbb{R}^N$  is a viscosity solution of (1) if:

- (1)  $u$  is a viscosity subsolution of (1), i.e., for every point  $(t_0, x_0) \in ]0, T[ \times \mathbb{R}^N$  and test function  $v \in C^1((0, +\infty) \times \mathbb{R}^N)$  such that  $u - v$  has a local maximum at  $(t_0, x_0)$ , it holds

$$v_t(t_0, x_0) + H(\nabla_x v(t_0, x_0)) \leq 0,$$

- (2)  $u$  is a viscosity supersolution of (1), i.e., for every point  $(t_0, x_0) \in ]0, T[ \times \mathbb{R}^N$  and test function  $v \in C^1((0, +\infty) \times \mathbb{R}^N)$  such that  $u - v$  has a local minimum at  $(t_0, x_0)$ , it holds

$$v_t(t_0, x_0) + H(\nabla_x v(t_0, x_0)) \geq 0.$$

In addition, we say that  $u$  is a viscosity solution of the Cauchy problem (1)-(2) if condition (2) is satisfied in the classical sense.

**Remark 3.** By the alternative equivalent definition of viscosity solution expressed in terms of the sub- and superdifferential of the function (see [11]), and because of Theorem 3-(iv), one immediately see that every  $C^1$  solution of (1) is also a viscosity solution of (1). On the other hand, if  $u$  is a viscosity solution of (1), then  $u$  satisfies the equation at every point of differentiability. Moreover, by the definition of reachable gradient, it follows that there holds

$$p_t + H(p_x) = 0 \quad \forall (p_t, p_x) \in D^*u(t, x), \quad (20)$$

at any  $(t, x) \in [0, +\infty[ \times \mathbb{R}^N$ .

It is well-known that a viscosity solution  $u$  of (1) is locally semiconcave (see, for instance, [9, Theorem 5.3.8]). Relying on the properties of the semiconcave functions recalled in the previous section, one can prove further regularity for viscosity solutions which will be useful in the paper.

**Proposition 3.** Let  $u : [0, T] \times \mathbb{R}^N$  be a viscosity solution of (1) and assume that  $u(t, \cdot)$  is both semiconcave and semiconvex in  $\mathbb{R}^N$  for all  $t \in ]0, T]$ . Then  $u$  is a continuously differentiable classical solution of (1) on  $]0, T] \times \mathbb{R}^N$ .

In other words, smoothness in the pair  $(t, x)$  follows from smoothness in the second variable. We give a proof for the reader's convenience.

*Proof.* Since a viscosity solution is locally semiconcave, relying on property (i) of Theorem 2 and properties (ii), (iv) of Theorem 3 it follows that, in order to show that  $u$  is everywhere continuously differentiable, it is sufficient to prove that the superdifferential  $D^+u(t, x)$  is a singleton for all  $(t, x) \in ]0, T] \times \mathbb{R}^N$ . In turn, the differentiability of  $u$  implies that the equation (1) is pointwise satisfied in the classical sense by Remark 3. Then, fix  $(t_0, x_0) \in ]0, T] \times \mathbb{R}^N$  and observe that, by Remark 2,  $u(t_0, \cdot)$  is differentiable at  $x_0$  since it is both semiconcave and semiconvex in  $\mathbb{R}^N$ . Therefore, by property (iv) of Theorem 3, the superdifferential  $D_x^+u(t_0, x_0)$  of  $u(t_0, \cdot)$  at  $x_0$  is the singleton  $\{\nabla_x u(t_0, x_0)\}$ . On the other hand, invoking a well-known property of the superdifferential (see, for instance, [9, Lemma 3.3.16]) we deduce that  $\Pi_x D^+u(t_0, x_0) = D_x^+u(t_0, x_0)$ , where  $\Pi_x$  denotes the projection of  $\mathbb{R} \times \mathbb{R}^N$  onto  $\mathbb{R}^N$  defined by  $\Pi_x(t, x) = x$ . Hence, recalling property (i) of Theorem 3, we get  $D^+u(t_0, x_0) = [\tau_-, \tau_+] \times \{\nabla_x u(t_0, x_0)\}$  for some  $\tau_{\pm} \in \mathbb{R}$ , with  $\tau_- \leq \tau_+$ . This implies that  $(\tau_{\pm}, \nabla_x u(t_0, x_0)) \in D^*u(t_0, x_0)$  by property (iii) of Theorem 3. So, applying (20), we find  $\tau_- + H(\nabla_x u(t_0, x_0)) = 0 = \tau_+ + H(\nabla_x u(t_0, x_0))$ , which in turn yields  $\tau_- = \tau_+$ , showing that  $D^+u(t_0, x_0)$  is a singleton as desired.  $\square$

Further analysis shows that, with the same hypotheses of Proposition 3, the viscosity solutions of (1) have a locally Lipschitz gradient in  $(t, x)$ .

Under assumption **(H1)'**, the viscosity solution of the Hamilton-Jacobi equation (1) with initial data  $u(0, \cdot) = u_0 \in \text{Lip}(\mathbb{R}^N)$  can be represented as the value function of a classical problem in calculus of variation, which admits the Hopf-Lax representation formula

$$u(t, x) = \min_{y \in \mathbb{R}^N} \left\{ t \cdot H^* \left( \frac{x - y}{t} \right) + u_0(y) \right\}, \quad t > 0, \quad x \in \mathbb{R}^N, \quad (21)$$

where  $H^*$  denotes the Legendre transform of  $H$ , defined by

$$H^*(q) := \max_{p \in \mathbb{R}^N} \{ p \cdot q - H(p) \} \quad q \in \mathbb{R}^N. \quad (22)$$

The Legendre transform inherits the properties of  $H$  (cfr. [9, Appendix A.2]). In particular, assumption **(H1)'** implies that  $H^* \in C^2(\mathbb{R}^N)$  and  $H^*$  is a uniformly convex coercive map, i.e.,

$$\lim_{|p| \rightarrow \infty} \frac{H^*(p)}{|p|} = +\infty. \quad (23)$$

Moreover,  $\nabla H^*$  is a  $C^1$  diffeomorphisms on  $\mathbb{R}^N$  as  $\nabla H$ , and one has

$$(\nabla H^*)^{-1}(p) = \nabla H(p), \quad D^2 H^*(p) = \left( D^2 H(\nabla H^*(p)) \right)^{-1} \quad \forall p \in \mathbb{R}^N. \quad (24)$$

On the other hand, the lower bound bound on the Hessian matrix  $D^2 H$  given in **(H1)** implies

$$D^2 H^* \leq \frac{1}{\alpha} \cdot \mathbb{I}_N, \quad (25)$$

while, by virtue of **(H2)**, we have

$$H^*(0) = -H(0). \quad (26)$$

The main properties of viscosity solutions defined by the Hopf-Lax formula of interest to this paper are recalled below (cfr. [9, Section 1.2, Section 6.4], [13, Section 3.3]).

**Proposition 4.** *Let  $u$  be the viscosity solution of (1) on  $[0, +\infty[ \times \mathbb{R}^N$ , with initial data  $u_0 \in \text{Lip}(\mathbb{R}^N)$ , defined by (21). Then the following holds true.*

(i) **Functional identity:** *for all  $x \in \mathbb{R}^N$  and  $0 \leq s < t$ , it holds*

$$u(t, x) = \min_{y \in \mathbb{R}^N} \left\{ u(s, y) + (t - s) \cdot H^* \left( \frac{x - y}{t - s} \right) \right\}.$$

(ii) **Differentiability of  $u$  and uniqueness:** *for all  $x \in \mathbb{R}^N$  and  $t > 0$ , any minimizer  $y_x$  of (21) satisfies  $y_x \in \{x - t \cdot \nabla H(p) \mid p \in D_x^*(u(t, x))\}$ , where  $D_x^*(u(t, x))$  denotes the reachable gradient of  $u(t, \cdot)$  at  $x$ . Moreover, (21) admits a unique minimizer  $y_x$  if and only if  $u(t, \cdot)$  is differentiable at  $x$ . In this case we have that  $y_x = x - t \cdot \nabla H(\nabla_x(u(t, x)))$ .*

(iii) **Dynamic programming principle:** *let  $t > s > 0$ ,  $x \in \mathbb{R}^N$ , assume that  $y$  is a minimizer for (21), and define  $z = \frac{s}{t}x + (1 - \frac{s}{t})y$ . Then  $y$  is the unique minimizer over  $\mathbb{R}^N$  of*

$$w \mapsto s \cdot H^* \left( \frac{z - w}{s} \right) + u_0(w) \quad (w \in \mathbb{R}^N).$$

By the above observations and because of Proposition 4-(i), the family of nonlinear operators

$$S_t : \text{Lip}(\mathbb{R}^N) \rightarrow \text{Lip}(\mathbb{R}^N), \quad u_0 \mapsto S_t u_0, \quad t \geq 0,$$

defined by

$$\begin{cases} S_t u_0(x) := \min_{y \in \mathbb{R}^N} \left\{ t \cdot H^*\left(\frac{x-y}{t}\right) + u_0(y) \right\} & t > 0, \quad x \in \mathbb{R}^N, \\ S_0 u_0(x) := u(x) & x \in \mathbb{R}^N, \end{cases} \quad (27)$$

enjoy the following properties:

- (i) for every  $u_0 \in \text{Lip}(\mathbb{R}^N)$ ,  $u(t, x) := S_t u_0(x)$  provides the unique viscosity solution of the Cauchy problem (1)-(2);
- (ii) (semigroup property)

$$S_{t+s} u_0 = S_t S_s u_0, \quad \forall t, s \geq 0, \quad \forall u_0 \in \text{Lip}(\mathbb{R}^N);$$

- (iii) for every constant  $c \in \mathbb{R}$  we have that

$$S_t(u_0 + c) = S_t u_0 + c, \quad \forall u_0 \in \text{Lip}(\mathbb{R}^N), \quad \forall t \geq 0. \quad (28)$$

It's a well-known fact that, for every fixed  $t \geq 0$ , the map  $S_t$  is continuous with respect to the topology of uniform convergence on compact sets. We next provide a proof of the continuity of such a map also in the case where the space  $\text{Lip}(\mathbb{R}^N)$  is endowed with the  $\mathbf{W}_{loc}^{1,1}$ -topology and  $S_t$  is restricted to sets of functions with uniform Lipschitz constant. Namely, the following holds.

**Proposition 5.** *Let  $u, u^\nu \in \text{Lip}(\mathbb{R}^N)$  ( $\nu \in \mathbb{N}$ ) be such that*

$$\text{Lip}[u^\nu] \leq M \quad \forall \nu, \quad \text{for some } M > 0, \quad (29)$$

$$u^\nu \xrightarrow{\nu \rightarrow \infty} u \quad \text{in} \quad \mathbf{W}_{loc}^{1,1}(\mathbb{R}^N). \quad (30)$$

*Then, for every fixed  $t \geq 0$ , one has*

$$S_t u^\nu \xrightarrow{\nu \rightarrow \infty} S_t u \quad \text{in} \quad \mathbf{W}_{loc}^{1,1}(\mathbb{R}^N). \quad (31)$$

*Proof.* In order to establish the proposition it will be sufficient to show that, for every given bounded domain  $\Omega \subset \mathbb{R}^N$ , and for any fixed  $t \geq 0$ , there holds

$$S_t u^\nu \xrightarrow{\nu \rightarrow \infty} S_t u \quad \text{in} \quad \mathbf{W}^{1,1}(\Omega). \quad (32)$$

Observe that, because of (29), and relying on the a-priori bound on the gradient of the solution to (1) provided by Lemma 1 in the next section, we have  $\text{Lip}[S_t u^\nu] \leq M$  for all  $\nu$ . In turn, this implies  $|p| \leq M$  for all  $p \in D_x^* S_t u^\nu(x)$ ,  $x \in \mathbb{R}^N$ , and for any  $\nu$ . Thus, invoking Proposition 4-(ii) we deduce that, for all  $x \in \Omega$  and for any minimizer  $y_x^\nu$  of (21), with  $u^\nu$  in place of  $u_0$ , one has

$$y_x^\nu \in \Omega' := \{x \in \mathbb{R}^N \mid d(x, \Omega) \leq t \cdot \sup_{|p| \leq M} |\nabla H(p)|\} \quad \forall \nu. \quad (33)$$

Next, notice that because of (29), (30), letting  $\bar{x} \in \Omega$  be a point such that  $u(\bar{x}) = \lim_{\nu \rightarrow \infty} u^\nu(\bar{x})$ , we have

$$\begin{aligned} |u^\nu(x)| &\leq \sup_\nu |u^\nu(\bar{x})| + M \cdot \text{diam}(\Omega') < +\infty \quad \forall x \in \overline{\Omega'}, \\ |u^\nu(x) - u^\nu(y)| &\leq M \cdot |x - y| \quad \forall x, y \in \overline{\Omega'}, \end{aligned} \quad \forall \nu. \quad (34)$$

Therefore, by a standard argument based on (30) and the Ascoli-Arzelà compactness theorem, we deduce that

$$u^\nu \xrightarrow{\nu \rightarrow \infty} u \quad \text{uniformly on } \Omega'. \quad (35)$$

Repeating the same reasoning for every bounded domain of  $\mathbb{R}^N$  it follows that  $u^\nu$  converges to  $u$  (uniformly on compact sets) on the whole space  $\mathbb{R}^N$  and that  $\text{Lip}[u] \leq M$ . Hence, for all  $x \in \Omega$  and for any minimizer  $y_x$  of (21), with  $u$  in place of  $u_0$ , one has  $y_x \in \Omega'$ . In turn, together with (33) and (35), this fact implies that

$$\min_{y \in \mathbb{R}^N} \left\{ t \cdot H^*\left(\frac{x-y}{t}\right) + u^\nu(y) \right\} \xrightarrow{\nu \rightarrow \infty} \min_{y \in \mathbb{R}^N} \left\{ t \cdot H^*\left(\frac{x-y}{t}\right) + u(y) \right\} \quad \text{uniformly on } \Omega, \quad (36)$$

which, by virtue of definition (27), yields

$$S_t u^\nu(x) \xrightarrow{\nu \rightarrow \infty} S_t u(x) \quad \text{uniformly on } \Omega. \quad (37)$$

As a consequence, we deduce that

$$S_t u^\nu \xrightarrow{\nu \rightarrow \infty} S_t u \quad \text{in } \mathbf{L}^1(\Omega). \quad (38)$$

On the other hand, observe that by Proposition 4-(ii) it follows that  $S_t u^\nu$ ,  $S_t u$  are differentiable almost everywhere in  $\Omega$ , and there holds

$$\begin{aligned} y_x^\nu &= x - t \cdot \nabla H(\nabla S_t u^\nu(x)) \quad \forall \nu, \\ y_x &= x - t \cdot \nabla H(\nabla S_t u(x)), \end{aligned} \quad \text{for a.e. } x \in \Omega, \quad (39)$$

where  $y_x^\nu$ ,  $y_x$  denotes the unique minimizer of (21), with  $u^\nu$  and  $u$  in place of  $u_0$ , respectively. Moreover, because of the uniqueness of such minimizers of (21), and by virtue of the convergence (36), (37), we deduce that  $\{y_x^\nu\}_\nu$  converges to  $y_x$  for almost every  $x \in \Omega$ . Thus, relying on (39), and recalling that  $\nabla H$  is a diffeomorphism on  $\mathbb{R}^N$ , we infer that

$$\nabla S_t u^\nu(x) \xrightarrow{\nu \rightarrow \infty} \nabla S_t u(x) \quad \text{for a.e. } x \in \Omega. \quad (40)$$

On the other hand, these unique minimizers satisfy  $y_x^\nu \in \Omega'$  for all  $\nu$  and for almost every  $x \in \Omega$ , so that one has

$$\left| \frac{x - y_x^\nu}{t} \right| \leq \sup_{|p| \leq M} |\nabla H(p)| \quad \forall \nu, \quad \text{for a.e. } x \in \Omega. \quad (41)$$

Thus, because of (39), (41), we derive a uniform  $\mathbf{L}^\infty$  bound on  $\nabla S_t u^\nu$ ,  $\nu \in \mathbb{N}$ , over  $\Omega$ , which, together with (40), implies

$$\nabla S_t u^\nu \xrightarrow{\nu \rightarrow \infty} \nabla S_t u \quad \text{in } \mathbf{L}^1(\Omega). \quad (42)$$

Then, from (38), (42) we recover (32), concluding the proof of the proposition.  $\square$

### 3 Upper estimates

#### 3.1 A-priori bounds on the Hopf-Lax semigroup

Let  $H : \mathbb{R}^N \rightarrow \mathbb{R}$  be a function satisfying the assumptions **(H1)**-**(H2)**. We collect here some a-priori bounds on the semiconcavity constant and on the gradient of the solutions to (1) and we establish an a-priori bound on the size of their support. Namely, given  $L, M > 0$ , consider the set of initial data introduced in (3):

$$\mathcal{C}_{[L,M]} = \left\{ u_0 \in \text{Lip}(\mathbb{R}^N) \mid \text{supp}(u_0) \subset [-L, L]^N, \text{Lip}[u_0] \leq M \right\}.$$

The image of  $\mathcal{C}_{[L,M]}$  through the Hopf-Lax semigroup map  $S_T$  defined in (27) enjoy the properties stated in the following

**Lemma 1.** *For any  $L, M, T > 0$  and for every  $u_0 \in \mathcal{C}_{[L,M]}$ , the following properties hold true:*

- (i)  $S_T u_0$  is semiconcave in  $\mathbb{R}^N$  with constant  $\frac{1}{\alpha T}$ ;
- (ii)  $\text{Lip}[S_T u_0] \leq M$ ;
- (iii)  $\text{supp}(S_T u_0 + T \cdot H(0)) \subset [-l_{[L,M,T]}, l_{[L,M,T]}]^N$ , where  $l_{[L,M,T]}$  is the constant defined in (6).

*Proof.* Under the assumption **(H1)** and recalling (25), property (i) is well-known (see [9, Corollary 1.6.2]), while  $\text{Lip}[u_0] \leq M$  and an application of [9, Theorem 1.3.2] implies

$$|S_T u_0(y) - S_T u_0(x)| \leq M |y - x|, \quad \forall x, y \in \mathbb{R}^N,$$

which yields (ii).

Concerning a proof of (iii), by the Lipschitz continuity of  $S_T u_0$  it will be sufficient to show that at every point  $x \in \mathbb{R}^N \setminus [-l_{[L,M,T]}, l_{[L,M,T]}]^N$  where  $S_T u_0$  is differentiable there holds

$$S_T u_0(x) = -T \cdot H(0). \quad (43)$$

Indeed, recalling (27) and invoking Proposition 4-(ii), we find that at every such point  $x$  one has

$$S_T u_0(x) = T \cdot H^*(\nabla H(\nabla S_T u_0(x))) + u_0(y_x), \quad (44)$$

where

$$y_x = x - T \cdot \nabla H(\nabla S_T u_0(x)). \quad (45)$$

Observe now that, relying on the property (ii) above established and recalling (6), we deduce that  $y_x \in \mathbb{R}^N \setminus [-L, L]^N$  for all  $x \in \mathbb{R}^N \setminus [-l_{[L,M,T]}, l_{[L,M,T]}]^N$ . This, in turn, implies

$$u_0(y_x) = 0, \quad \nabla u_0(y_x) = 0, \quad (46)$$

because  $\text{supp}(u_0) \subset [-L, L]^N$  by the definition (3) of the set  $\mathcal{C}_{[L,M]}$ . Moreover, since by Proposition 4-(ii)  $y_x$  is a minimum of

$$y \mapsto T \cdot H^*\left(\frac{x - y}{T}\right) + u_0(y)$$

over  $\mathbb{R}^N$ , it follows that  $-\nabla H^*\left(\frac{x - y_x}{T}\right) = \nabla u_0(y_x)$ . Hence, relying on (45), (46), we deduce that

$$\nabla H^*(\nabla H(\nabla S_T u_0(x))) = \nabla H^*\left(\frac{x - y_x}{T}\right) = 0. \quad (47)$$

Thus, by virtue of (44), (46), (47), we conclude that  $\nabla S_T(u)(x) = 0$  at every point  $x \in \mathbb{R}^N \setminus [-l_{[L,M,T]}, l_{[L,M,T]}]^N$  where  $S_T u_0$  is differentiable. This, in turn, by the assumption **(H2)** and because of (44), (46), implies that at every such point  $x$  there holds

$$S_T u_0(x) = T \cdot H^*(\nabla H(0)) = T \cdot H^*(0). \quad (48)$$

Finally, recalling (26), we recover (43) from (48), thus completing the proof of (iii).  $\square$

**Remark 4.** Property (iii) of Lemma 1 implies that, for every  $u_0 \in \mathcal{C}_{[L,M]}$ , the domain  $\text{supp}(S_t u_0 + t \cdot H(0))$  where  $S_t u_0$  differs from the constant in space solution with zero initial data propagates at a finite speed as illustrated in Figure 1 below.

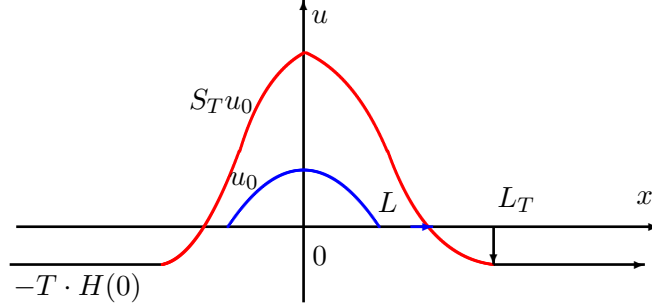


Figure 1 : evolution of the support of the gradient under  $S_T$ .

Having in mind the a-priori bound established in [12] for the support of solutions to scalar conservation laws with convex flux, one may wonder whether it is possible to derive a sharper estimate on the size of such a domain. In fact, if we consider a class of initial data

$$\mathcal{C}_{[L,m,M]} := \left\{ u_0 \in \text{Lip}(\mathbb{R}^N) \mid \text{supp}(u_0) \subset [-L, L]^N, \|\nabla u_0\|_{\mathbf{L}^1(\mathbb{R}^N)} \leq m, \text{Lip}[u_0] \leq M \right\}, \quad (49)$$

one may look for establishing an estimate as

$$|\text{supp}(S_T u_0 + T \cdot H(0))| \leq \left( 2L + \sup_{|p| \leq M} \|D^2 H(p)\| \cdot 4 \sqrt{\frac{mT}{\alpha}} \right)^N \quad (50)$$

relying on property (i) of Lemma 1 and property (iii) of Theorem 2. However, a key point in the proof of an estimate of this type for the support of solutions to scalar conservation laws is the fact that, for such equations, the  $\mathbf{L}^1$ -norm of the solution is non increasing in time as a consequence of the  $\mathbf{L}^1$  contractivity of the semigroup map  $S_t$ . This property continues to hold for the gradient of solutions to Hamilton-Jacobi equations in one space dimension, but it is no more true in general when the space dimension is greater than one. In fact in this case, as observed in the introduction, the gradient of a solution of an Hamilton-Jacobi equation turns out to be a solution of an hyperbolic system of conservation laws and it is well-known that for general hyperbolic systems of conservation laws no metric is contractive [20]. As a consequence, one can easily convince himself that a bound as (50) doesn't hold for Hamilton-Jacobi equations in several space variables. This is the main reason for which we limit ourself to analyze in this paper the image through the Hopf-Lax semigroup  $S_t$  of sets of initial data of the form (3) and we don't consider sets of the form (49).

Given any  $L, M, K > 0$ , consider now the class of functions

$$\mathcal{SC}_{[L,M,K]} := \left\{ u \in \mathcal{C}_{[L,M]} \mid u \text{ is semiconcave with semiconcavity constant } K \right\}, \quad (51)$$

where  $\mathcal{C}_{[L,M]}$  denotes the set in (3). Then, applying Lemma 1, we immediately obtain the following.

**Proposition 6.** *Let  $H : \mathbb{R}^N \rightarrow \mathbb{R}$  be a function satisfying the assumptions **(H1)**-**(H2)** and  $\{S_t : \text{Lip}(\mathbb{R}^N) \rightarrow \text{Lip}(\mathbb{R}^N)\}_{t \geq 0}$  be the semigroup of viscosity solutions generated by (1). Then, given any  $L, M, T > 0$ , there holds*

$$S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0) \subset \mathcal{SC}_{[l_{[L,M,T]}, M, \frac{1}{\alpha T}]}, \quad (52)$$

where  $l_{[L,M,T]}$  is given by (6) and  $\alpha$  is the constant in **(H1)**.

### 3.2 An upper bound on the $\varepsilon$ -entropy for semiconcave functions

Towards a derivation of an upper bound on the  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  for the class of semiconcave functions introduced in (51), in view of Proposition 2-(ii) we shall first establish an upper bound on the  $\varepsilon$ -entropy in  $\mathbf{L}^1$  for a class of monotone multifunctions with uniformly bounded total variation defined on a cube of  $\mathbb{R}^N$ . As observed in Section 2.1, any monotone multifunction is almost everywhere univalued in the interior of its domain, and can be regarded as a function of bounded variation on this set. Hence, set  $I_L := ] - L, L[$ ,  $I_M := ] - M, M[$ , and consider the class of monotone multifunction

$$\mathcal{F}_{[L,M,C]} := \left\{ F : I_L^N \rightarrow 2^{I_M^N} \mid \text{dom}(F) = I_L^N, \ F \text{ is decreasing, } |DF|(I_L^N) \leq C \right\}, \quad (53)$$

where  $|DF|$  denotes the total variation of the matrix-valued Radon measure  $DF$ . With a slight abuse of notation, we shall regard  $\mathcal{F}_{[L,M,C]}$  as a subset of  $\mathbf{L}^1(I_L^N, I_M^N)$  consisting of all functions in  $\mathbf{L}^1(I_L^N, I_M^N)$  that coincide almost everywhere with an element of the set defined in (53).

**Proposition 7.** *Given  $L, M > 0$ , for any  $\varepsilon > 0$  sufficiently small there holds*

$$\mathcal{H}_\varepsilon(\mathcal{F}_{[L,M,C]} \mid \mathbf{L}^1(I_L^N, I_M^N)) \leq \gamma_{[L,M,C,N]} \cdot \frac{1}{\varepsilon^N}, \quad (54)$$

where

$$\gamma_{[L,M,C,N]} := 2^{(N^2+N+1)} N^{\left(\frac{N}{2}+1\right)} \cdot (ML^N + LC)^N. \quad (55)$$

*Proof.*

**1.** Towards a proof of (54), we shall associate to any function  $F \in \mathcal{F}_{[L,M,C]}$ , a piecewise constant function  $\tilde{F} \in \mathbf{L}^\infty(I_L^N, I_M^N)$  that takes values in a discrete subset of  $I_M^N$  and has the property that every  $i$ -th component is (almost everywhere) monotone decreasing in the  $i$ -th variable. Namely, given any fixed  $n \in \mathbb{N}$ , we divide  $] - L, L[$  into  $n^N$  cubes with sides of length  $\frac{2L}{n}$  as follows. For every multiindex  $\iota = (\iota_1, \dots, \iota_N) \in \{0, \dots, n-1\}^N$  we define the cube

$$\square_\iota := \left] -L + \frac{\iota_1}{n} 2L, -L + \frac{\iota_1 + 1}{n} 2L \right[ \times \cdots \times \left] -L + \frac{\iota_N}{n} 2L, -L + \frac{\iota_N + 1}{n} 2L \right[ ,$$

so that one has

$$\bar{I}_L^N = \bigcup_{\iota \in \{0, \dots, n-1\}^N} \bar{\square}_\iota, \quad (56)$$

where  $\bar{I}_L = [-L, L]$ , and  $\bar{\square}_\iota$  denotes the closure of  $\square_\iota$ . Then, given any  $F \in \mathcal{F}_{[L,M,C]}$ , for every  $\iota \in \{0, \dots, n-1\}^N$ , let

$$\bar{F}_\iota := \frac{1}{\text{Vol}(\square_\iota)} \cdot \int_{\square_\iota} F(x) dx \quad (57)$$

be the average of  $F$  over  $\square_\iota$ . Observe that  $\bar{F}_\iota = (\bar{F}_\iota^1, \dots, \bar{F}_\iota^N) \in I_M^N$  since  $F$  takes values in  $I_M^N$ . Next, consider the subdivision of  $[-M, M]$  into the  $n$  intervals

$$[-M, M] = \left[ -M, -M + \frac{2M}{n} \right] \cup \dots \cup \left[ -M + \frac{2(n-2)}{n}M, -M + \frac{2(n-1)}{n}M \right] \cup \left[ -M + \frac{2(n-1)}{n}M, M \right],$$

and define the vector  $\tilde{F}_\iota = (\tilde{F}_\iota^1, \dots, \tilde{F}_\iota^N) \in I_M^N$  by setting, for each  $i \in \{1, \dots, N\}$ ,

$$\tilde{F}_\iota^i := \begin{cases} -M + (k + \frac{1}{2}) \cdot \frac{2M}{n}, & \text{if } \bar{F}_\iota^i \in \left[ -M + 2k \cdot \frac{M}{n}, -M + 2(k+1) \cdot \frac{M}{n} \right], & k \leq n-2, \\ -M + (n - \frac{1}{2}) \cdot \frac{2M}{n}, & \text{if } \bar{F}_\iota^i \in \left[ -M + (n-1) \cdot \frac{M}{n}, M \right]. \end{cases} \quad (58)$$

Then, let  $\tilde{F} \in \mathbf{L}^\infty(I_L^N, I_M^N)$  be the function defined almost everywhere by

$$\tilde{F}(x) := \tilde{F}_\iota \quad \text{if } x \in \square_\iota, \quad \iota \in \{0, \dots, n-1\}^N, \quad (59)$$

with  $\tilde{F}_\iota = (\tilde{F}_\iota^1, \dots, \tilde{F}_\iota^N)$  as in (58). By construction, we have

$$\tilde{F}(x) \in J_{M,n} := \left\{ -M + \left(k + \frac{1}{2}\right) \cdot \frac{2M}{n} \mid k = 0, \dots, n-1 \right\} \quad \forall x \in \bigcup_\iota \square_\iota. \quad (60)$$

We claim that  $F$  enjoys the following two properties:

(i) For every  $\iota = (\iota_1, \dots, \iota_N) \in \{0, \dots, n-1\}^N$  there holds

$$\iota_i < n-1 \implies \tilde{F}_\iota^i \geq \tilde{F}_{\iota+e_i}^i, \quad (61)$$

where  $e_i$  denotes the  $i$ -th element of the canonical basis of  $\mathbb{R}^N$ .

(ii)

$$\|\tilde{F} - F\|_{\mathbf{L}^1(I_L^N, I_M^N)} \leq \gamma_{[L,M,C,N]}^1 \cdot \frac{1}{n} \quad (62)$$

where

$$\gamma_{[L,M,C,N]}^1 := \sqrt{N} \left( M(2L)^N + LC \right). \quad (63)$$

In fact, given any  $\iota = (\iota_1, \dots, \iota_N) \in \{0, \dots, n-1\}^N$  and  $\iota_i < n-1$ , by definition (57) and since  $F$  is monotone decreasing we find

$$\begin{aligned} \bar{F}_{\iota+e_i}^i - \bar{F}_\iota^i &= \langle \bar{F}_{\iota+e_i} - \bar{F}_\iota, e_i \rangle \\ &= \left\langle \frac{n^N}{(2L)^N} \cdot \int_{\square_\iota} \left( F\left(x + \frac{2L}{n}e_i\right) - F(x) \right) dx, e_i \right\rangle \\ &= \frac{n^{N+1}}{(2L)^{N+1}} \cdot \int_{\square_\iota} \left\langle F\left(x + \frac{2L}{n}e_i\right) - F(x), x + \frac{2L}{n}e_i - x \right\rangle dx \\ &\leq 0. \end{aligned} \quad (64)$$

By definition (58), (64) implies, in turn,  $\tilde{F}_{\iota+e_i}^i \leq \tilde{F}_\iota^i$ , thus proving (61). Concerning (62), observe first that, by definition (57) and relying on the Poincaré inequality for BV functions stated in Theorem 4, for any  $\iota = (\iota_1, \dots, \iota_N) \in \{0, \dots, n-1\}^N$  we derive

$$\|F - \bar{F}_\iota\|_{\mathbf{L}^1(\square_\iota, I_M^N)} \leq \frac{L\sqrt{N}}{n} \cdot |DF|(\square_\iota). \quad (65)$$



On the other hand, since (58) implies

$$|\tilde{F}_\iota - \bar{F}_\iota| \leq \frac{\sqrt{NM}}{n},$$

it follows that

$$\|\tilde{F} - \bar{F}_\iota\|_{\mathbf{L}^1(\square_\iota, I_M^N)} \leq \frac{\sqrt{NM} \cdot (2L)^N}{n^{N+1}}. \quad (66)$$

Combining the estimates (65), (66), and observing that, by definition (53),  $F \in \mathcal{F}_{[L,M,C]}$  implies  $|DF|(I_L^N) \leq C$ , we obtain

$$\begin{aligned} \|\tilde{F} - F\|_{\mathbf{L}^1(I_L^N, I_M^N)} &\leq \sum_\iota \left( \|\tilde{F} - \bar{F}_\iota\|_{\mathbf{L}^1(\square_\iota, I_M^N)} + \|F - \bar{F}_\iota\|_{\mathbf{L}^1(\square_\iota, I_M^N)} \right) \\ &\leq \frac{\sqrt{NM} \cdot (2L)^N}{n} + \frac{L\sqrt{N}}{n} \cdot |DF|(I_L^N) \\ &\leq \frac{\sqrt{N}}{n} (M(2L)^N + LC), \end{aligned} \quad (67)$$

proving (62).

**2.** We introduce now a set of piecewise constant functions sharing the properties (60), (61) of  $\tilde{F}$ . Namely, letting  $J_{M,n}$  be the set in (60), we define

$$\begin{aligned} \mathcal{G}_n := \left\{ G \in \mathbf{L}^\infty(I_L^N, J_{M,n}^N) \mid G \text{ is constant on every } \square_\iota, \iota = (\iota_1, \dots, \iota_N) \in \{0, \dots, n-1\}^N \right. \\ \left. \text{and } \iota_i < n-1 \implies G_\iota^i \geq G_{\iota+e_i}^i \right\}, \end{aligned} \quad (68)$$

where  $G_\iota$  stands for the value of  $G$  on the cube  $\square_\iota$ . Observe that for every  $F \in \mathcal{F}_{[L,M,C]}$ , letting  $\tilde{F}$  be the map defined in (59), one has  $\tilde{F} \in \mathcal{G}_n$ . Moreover, setting for any given  $G \in \mathcal{G}_n$

$$U(G) := \left\{ F \in \mathbf{L}^1(I_L^N, I_M^N) \mid \|F - G\|_{\mathbf{L}^1} \leq \gamma_{[L,M,C,N]}^1 \cdot \frac{1}{n} \right\}, \quad (69)$$

with  $\gamma_{[L,M,C,N]}^1$  as in (63), because of (62) we have that  $F \in \mathcal{F}_{[L,M,C]}$  implies  $F \in U(\tilde{F})$ . Hence, the set

$$\mathcal{U} := \left\{ U(G) \mid G \in \mathcal{G}_n \right\}$$

provides an  $\mathbf{L}^1$  covering of  $\mathcal{F}_{[L,M,C]}$  with sets of diameter  $2\gamma_{[L,M,C,N]}^1 \cdot \frac{1}{n}$ . Thus, taking

$$n = \left\lfloor \frac{\gamma_{[L,M,C,N]}^1}{\varepsilon} \right\rfloor + 1, \quad (70)$$

we deduce that

$$\mathcal{N}_\varepsilon(\mathcal{F}_{[L,M,C]} \mid \mathbf{L}^1(I_L^N, I_M^N)) \leq \text{Card}(\mathcal{G}_n). \quad (71)$$

Observe that, given any fixed  $\bar{i} \in \{1, \dots, N\}$ , the set of piecewise constant scalar functions

$$\begin{aligned} \mathcal{G}_n^{\bar{i}} := \left\{ g \in \mathbf{L}^\infty(I_L^N, J_{M,n}) \mid g \text{ is constant on every } \square_\iota, \iota = (\iota_1, \dots, \iota_N) \in \{0, \dots, n-1\}^N \right. \\ \left. \text{and } \iota_{\bar{i}} < n-1 \implies g_\iota \geq g_{\iota+e_{\bar{i}}} \right\}, \end{aligned} \quad (72)$$

defined with the same notations as in (68), is independent of the choice of  $\bar{i}$ . Thus, we deduce from (71) that there holds

$$\mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,M,C]} \mid \mathbf{L}^1(I_L^N, I_M^N)\right) \leq (\text{Card}(\mathcal{G}_n^{\bar{i}}))^N. \quad (73)$$

Next, we define the set

$$\mathcal{J} := \left\{ \iota \in \{0, \dots, n-1\}^N \mid \iota_{\bar{i}} = -M \right\},$$

which collects all the labels of squares  $\square_\iota$  with boundary intersecting the hyperplane  $x_{\bar{i}} = -M$ . Consider the set of decreasing  $n$ -tuples of elements of the set  $J_{M,n}$  in (60)

$$\mathcal{K} := \left\{ (a_0, a_1, \dots, a_{n-1}) \in (J_{M,n})^n \mid -M + \left(n - \frac{1}{2}\right) \cdot \frac{2M}{n} \geq a_0 \geq \dots \geq a_{n-1} \geq -M + \frac{M}{n} \right\}.$$

By the definition (72) we deduce that

$$\text{Card}(\mathcal{G}_n^{\bar{i}}) \leq (\text{Card}(\mathcal{K}))^{\text{Card}(\mathcal{J})}. \quad (74)$$

Observe that  $\mathcal{K}$  has the same cardinality as the set of decreasing  $n$ -tuples of nonnegative integers smaller than  $n-1$ . By elementary combinatorial arguments it thus follows that, if  $n \geq 6$ , one has  $\text{Card}(\mathcal{K}) \leq \binom{2n}{n} \leq 2^{2n}$  (e.g. see [12, proof of Lemma 3.1]). Therefore, since  $\text{Card}(\mathcal{J}) = n^{N-1}$ , we derive from (73), (74) the upper bound

$$\mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,M,C]} \mid \mathbf{L}^1(I_L^N, I_M^N)\right) \leq (2^{2n})^{(Nn^{N-1})} = 2^{2Nn^N}. \quad (75)$$

Then, for every  $0 < \varepsilon \leq \frac{1}{5} \cdot \gamma_{[L,M,C,N]}^1$ , with  $\gamma_{[L,M,C,N]}^1$  as in (63), taking  $n$  as in (70) we recover from (75) the estimate

$$\mathcal{N}_\varepsilon\left(\mathcal{F}_{[L,M,C]} \mid \mathbf{L}^1(I_L^N, I_M^N)\right) \leq 2^{\frac{\gamma_{[L,M,C,N]}}{\varepsilon^N}}, \quad (76)$$

where  $\gamma_{[L,M,C,N]} := 2N(2 \cdot \gamma_{[L,M,C,N]}^1)^N$ . In turn, (76) yields (54), completing the proof of the proposition.  $\square$

Relying on Proposition 7 we now establish an upper bound on the  $\varepsilon$ -entropy in  $\mathbf{W}^{1,1}$  for the class of semiconcave functions introduced in (51).

**Proposition 8.** *Given  $L, M, K > 0$ , let  $\mathcal{SC}_{[L,M,K]}$  be the set defined in (51). Then, for  $\varepsilon > 0$  sufficiently small, there holds*

$$\mathcal{H}_\varepsilon\left(\mathcal{SC}_{[L,M,K]} \mid \mathbf{W}^{1,1}(\mathbb{R}^N)\right) \leq \gamma_{[L,M,K,N]}^{\mathcal{SC}} \cdot \frac{1}{\varepsilon^N} \quad (77)$$

where

$$\gamma_{[L,M,K,N]}^{\mathcal{SC}} := \omega_N^N \cdot \left(4N \cdot (1 + M + (K+1)L)\right)^{4N^2}. \quad (78)$$

*Proof.* Given  $L, K > 0$ , let us define the map  $\mathcal{T}_K : \mathbf{L}^1(I_L^N) \rightarrow \mathbf{L}^1(\mathbb{R}^N)$ ,  $I_L := ]-L, L[$ , that associates to any  $f \in \mathbf{L}^1(I_L^N)$  the function

$$\mathcal{T}_K f(x) := \begin{cases} f(x) + \frac{K}{2}|x|^2 & \text{if } x \in I_L^N, \\ 0 & \text{otherwise,} \end{cases} \quad (79)$$

and then consider the class of concave functions

$$\mathcal{C}_{[L,M,K]} := \left\{ f \in \mathbf{W}^{1,1}(I_L^N) \mid \mathcal{T}_K f \in \mathcal{SC}_{[L,M,K]} \right\}. \quad (80)$$

The definition (80) must be understood in the sense that a function  $f \in \mathbf{W}^{1,1}(I_L^N)$  is an element of  $\mathcal{C}_{[L,M,K]}$  if there exists  $\bar{f}$  almost everywhere equal to  $f$  such that  $\mathcal{T}_K \bar{f} \in \mathcal{SC}_{[L,M,K]}$ . Notice that, for any  $g \in \mathcal{SC}_{[K,L,M]}$ , letting  $\tilde{g} := (f - \frac{K}{2}|\cdot|^2)|_{I_L^N}$  denote the restriction of  $f - \frac{K}{2}|\cdot|^2$  to  $I_L^N$ , recalling definitions (3), (51) one has  $\tilde{g} \in \mathbf{W}^{1,1}(I_L^N)$  and  $\mathcal{T}_K \tilde{g} = g$ . Thus,  $\mathcal{T}_K$  is a surjective map from  $\mathcal{C}_{[L,M,K]}$  into  $\mathcal{SC}_{[K,L,M]}$ , and hence every given  $\varepsilon$ -covering  $\mathcal{B} = \{B_\alpha\}_\alpha$  of  $\mathcal{C}_{[K,L,M]}$  in  $\mathbf{W}^{1,1}(I_L^N)$  yields an  $\varepsilon$ -covering  $\{\mathcal{T}_K(B_\alpha)\}_\alpha$  of  $\mathcal{SC}_{[K,L,M]}$  in  $\mathbf{W}^{1,1}(\mathbb{R}^N)$  with the same cardinality. This implies that

$$\mathcal{H}_\varepsilon(\mathcal{SC}_{[L,M,K]} \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) \leq \mathcal{H}_\varepsilon(\mathcal{C}_{[L,M,K]} \mid \mathbf{W}^{1,1}(I_L^N)). \quad (81)$$

Therefore, in order to establish (77), it will be sufficient to show

$$\mathcal{H}_\varepsilon(\mathcal{C}_{[L,M,K]} \mid \mathbf{W}^{1,1}(I_L^N)) \leq \gamma_{[L,M,K,N]}^{\mathcal{SC}} \cdot \frac{1}{\varepsilon^N}. \quad (82)$$

**1.** Towards a proof of (82) observe that, for any given  $f \in \mathcal{C}_{[L,M,K]}$ , by definitions (3), (51), (80), and applying Proposition 2, there is a representative of  $f$ , that we still denote  $f$ , so that

- (i) the map  $x \mapsto f(x) + \frac{K}{2}|x|^2$  is semiconcave in  $I_L^N$  with constant  $K$  and has zero trace on  $\partial I_L^N$ ;
- (ii) the superdifferential  $D^+ f$  is a monotone decreasing multifunction in  $I_L^N$ ;
- (iii) there holds

$$\|\nabla f\|_{\mathbf{L}^\infty(I_L^N)} \leq M_1, \quad (83)$$

where

$$M_1 := M + K\sqrt{N}L. \quad (84)$$

By Theorem 3-(iii), in turn (83) yields

$$\text{diam}(D^+ f(I_L^N)) \leq 2M_1. \quad (85)$$

Then, relying on (ii) and on (85), and invoking Proposition 1, we obtain

$$|D^2 f|(I_L^N) \leq C_1, \quad (86)$$

with

$$C_1 := 2^{\frac{3N}{2}} \cdot N^{\left(\frac{N}{2}+2\right)} \cdot \omega_N \cdot (M + (K+1)L)^N, \quad (87)$$

where  $|D^2 f|$  denotes the total variation of the (matrix-valued) distributional derivative  $D^2 f$ . Therefore, if we consider the class of monotone multifunctions

$$\mathcal{DC}_{[L,M,K]} := \left\{ D^+ f \mid f \in \mathcal{C}_{[L,M,K]} \right\}, \quad (88)$$

recalling definition (53), by (83), (86) we have

$$\mathcal{DC}_{[L,M,K]} \subset \mathcal{F}_{[L,M_1,C_1]}, \quad (89)$$

and hence there holds

$$\mathcal{H}_\varepsilon\left(\mathcal{DC}_{[L,M,K]} \mid \mathbf{L}^1(I_L^N, I_{M_1}^N)\right) \leq \mathcal{H}_\varepsilon\left(\mathcal{F}_{[L,M_1,C_1]} \mid \mathbf{L}^1(I_L^N, I_{M_1}^N)\right), \quad (90)$$

with  $I_{M_1} := ]-M_1, M_1[$ . Thus, relying on Proposition 7, we find

$$\mathcal{H}_\varepsilon\left(\mathcal{DC}_{[L,M,K]} \mid \mathbf{L}^1(I_L^N, I_{M_1}^N)\right) \leq \gamma_{[L,M_1,C_1,N]} \cdot \frac{1}{\varepsilon^N} \quad (91)$$

where  $\gamma_{[L,M_1,C_1,N]}$  is a constant defined as in (55) with  $M_1, C_1$  given in (84), (87).

**2.** Relying on (91) and invoking the Poincaré inequality for trace-zero  $\mathbf{W}^{1,1}$  functions stated in Section 2.1, we shall produce now an  $\varepsilon$ -covering of  $\mathcal{C}_{[L,M,K]}$  in  $\mathbf{W}^{1,1}$  with a cardinality of order  $\leq \frac{(LN)^N \gamma_{[L,M_1,C_1,N]}}{\varepsilon^N}$ . In fact, observe that by property (i) above, for every  $f_1, f_2 \in \mathcal{C}_{[L,M,K]}$ , one has  $f_1 - f_2 \in \mathbf{W}_0^{1,1}(I_L^N)$ . Hence, applying the Poincaré inequality for  $\mathbf{W}_0^{1,1}$  functions stated in Theorem 4, we get

$$\|f_2 - f_1\|_{\mathbf{W}^{1,1}(I_L^N)} \leq (2L+1) \cdot \|\nabla f_2 - \nabla f_1\|_{\mathbf{L}^1(I_L^N, I_{M_1}^N)} \quad \forall f_1, f_2 \in \mathcal{C}_{[L,M,K]},$$

so that, for any  $f_1, f_2 \in \mathcal{C}_{[L,M,K]}$ , there holds

$$\|\nabla f_2 - \nabla f_1\|_{\mathbf{L}^1(I_L^N, I_{M_1}^N)} \leq \frac{\varepsilon}{(2L+1)} \implies \|f_2 - f_1\|_{\mathbf{W}_0^{1,1}(I_L^N)} \leq \varepsilon. \quad (92)$$

Next, by virtue of the estimate (91) on the  $\mathcal{H}_{\varepsilon'}$  entropy of  $\mathcal{DC}_{[L,M,K]}$  with  $\varepsilon' = \frac{\varepsilon}{2L+1}$ , there exist  $p$  functions  $f_1, f_2, \dots, f_p \in \mathcal{C}_{[L,M,K]}$ , with

$$p \leq \left\lfloor 2^{\left(\gamma_{[L,M_1,C_1,N]} \cdot \left(\frac{2L+1}{\varepsilon}\right)^N\right)} \right\rfloor, \quad (93)$$

so that

$$\mathcal{DC}_{[L,M,K]} \subset \bigcup_{l=1}^p B\left(D^+ f_l, \frac{\varepsilon}{(2L+1)}\right)$$

where  $B(D^+ f_l, \varepsilon)$  denotes the  $\mathbf{L}^1(I_L^N, I_{M_1}^N)$ -ball centered at  $D^+ f_l = \nabla f_l$  (regarded as an element of  $\mathbf{L}^1(I_L^N, I_{M_1}^N)$ ). Therefore, by definition (88) and because of (92), we deduce that

$$\mathcal{C}_{[L,M,K]} \subset \bigcup_{l=1}^p B(f_l, \varepsilon) \quad (94)$$

where  $B(f_l, \varepsilon)$  denotes the  $\mathbf{W}^{1,1}(I_L^N)$ -ball centered at  $f_l$ . Hence, observing that by (55), (84), (87), one has

$$\begin{aligned} (2L+1)^N \cdot \gamma_{[L,M_1,C_1,N]} &= (2L+1)^N \cdot 2^{(N^2+N+1)} N^{\left(\frac{N}{2}+1\right)} \cdot (M_1 L^N + L C_1)^N \\ &\leq (L+1)^N \cdot 2^{(N+1)^2} \cdot N^{N(N+3)} \cdot \omega_N^N \cdot ((M+KL)L^N + L(M+(K+1)L)^N)^N \\ &\leq 2^{(N+1)^2} \cdot N^{N(N+3)} \cdot \omega_N^N \cdot \left(1+M+(K+2)L\right)^{N(N+2)} \\ &\leq \omega_N^N \cdot \left(4N \cdot (1+M+(K+1)L)\right)^{4N^2}, \end{aligned} \quad (95)$$

it follows from (93), (94) that there holds (82) with  $\gamma_{[L,M,K,N]}^{SC}$  as in (78), thus completing the proof.  $\square$

### 3.3 Conclusion of the proof of Theorem 1-(i)

Given,  $L, M, T > 0$ , combining Proposition 6 and Proposition 8 we find that, for  $\varepsilon$  sufficiently small, there holds

$$\mathcal{H}_\varepsilon\left(S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0) \mid \mathbf{W}^{1,1}(\mathbb{R}^N)\right) \leq \gamma_{[l_{[L,M,T]}, M, \frac{1}{\alpha T}, N]}^{SC} \cdot \frac{1}{\varepsilon^N}, \quad (96)$$

where

$$\gamma_{[l_{[L,M,T]}, M, \frac{1}{\alpha T}, N]}^{SC} = \omega_N^N \cdot \left(4N \cdot \left(1 + M + (1/(\alpha T) + 1) \cdot l_{[L,M,T]}\right)\right)^{4N^2}, \quad (97)$$

with  $l_{[L,M,T]}$  as in (6). This establishes the upper bound (4).  $\square$

## 4 Lower estimates

### 4.1 Part 1: Controllability

Towards a proof of Theorem 1-(ii), we shall first show that, at every given time  $T > 0$ , one can represent the semiconcave functions of the set (51) as the values at time  $T$  of the Hopf-Lax solutions to (1) with initial data varying in a set of the form (3) translated by  $T \cdot H(0)$ , provided that the semiconcavity constant is sufficiently small.

**Proposition 9.** *Let  $H : \mathbb{R}^N \rightarrow \mathbb{R}$  be a function satisfying the assumptions **(H1)**-**(H2)** and  $\{S_t : \text{Lip}(\mathbb{R}^N) \rightarrow \text{Lip}(\mathbb{R}^N)\}_{t \geq 0}$  be the semigroup of viscosity solutions generated by (1). Then, given any  $L, M, T > 0$ , for every  $m, K > 0$  such that*

$$m \leq \min \left\{ m_0, M, \frac{L}{4 \|D^2 H(0)\| \cdot T} \right\}, \quad K \leq \frac{1}{4 \|D^2 H(0)\| \cdot T}, \quad (98)$$

where  $m_0$  is the constant in (19), there holds

$$\mathcal{SC}_{[L/2, m, K]} \subset S_T(\mathcal{C}_{[L, M]}) + T \cdot H(0), \quad (99)$$

where  $\mathcal{SC}_{[L/2, m, K]}$ ,  $\mathcal{C}_{[L, M]}$  denote sets defined as in (51), (3), respectively.

The proof of Proposition 9 is based on the lemma below, which shows that a solution of (1) with a semiconvex initial condition preserves the semiconvexity on a given time interval, provided the semiconvexity constant of the initial data is sufficiently small in absolute value.

**Lemma 2.** *In the same setting of Proposition 9, given  $M, T > 0$ , let  $u_0$  be a semiconvex function with semiconcavity constant  $-K$ . Assume that  $K > 0$  satisfies*

$$K \leq \frac{1}{2\alpha_M T} \quad \text{where} \quad \alpha_M := \sup_{|p| \leq M} \|D^2 H(p)\|, \quad (100)$$

and  $\text{Lip}[u_0] \leq M$ . Then, the following hold true.

- (i)  $x \mapsto S_t u_0(x)$  is semiconvex for all  $t \in [0, T]$ .
- (ii)  $(t, x) \mapsto S_t u_0(x)$  is a  $C^1$  classical solution of (1) on  $]0, T] \times \mathbb{R}^N$ .

*Proof of Proposition 9.* We will show that any element  $\psi$  of the set on the left-hand side of (99) can be obtained as the value at time  $T$  of a classical solution to (1) by reversing the direction of time, and constructing a backward solution to (1) that starts at time  $T$  from  $\psi$ . Namely, given

$$\psi \in \mathcal{SC}_{[L/2, m, K]}, \quad (101)$$

set

$$w_0(x) := -\psi(-x) \quad \forall x \in \mathbb{R}^N, \quad (102)$$

and consider the viscosity solution  $S_t w_0(x)$  of (1). Because of (101), (102), and by definitions (3), (51), we have  $w_0 \in \mathcal{C}_{[L/2, m]}$ . Moreover, recalling (6), (19), thanks to (98), (100), and to the assumption **(H2)**, one has

$$l_{[L/2, m, T]} \leq L/2 + T \cdot \sup_{|p| \leq m_0} \|D^2 H(p)\| \cdot m \leq L. \quad (103)$$

Hence, applying Lemma 1, we find

$$S_T w_0 \in \text{Lip}(\mathbb{R}^N), \quad \text{Lip}[S_T w_0] \leq m, \quad (104)$$

$$S_T w_0(x) = -T \cdot H(0) \quad \forall x \in \mathbb{R}^N \setminus [-L, L]^N. \quad (105)$$

On the other hand, notice that by (19), (51), (98), (101)  $\psi$  is a semiconcave function with semiconcavity constant  $K$  satisfying (100), with  $m$  in place of  $M$ . Then, it follows from (102) that  $w_0$  is semiconvex with semiconvexity constant  $-K$ . Thus, applying Lemma 2, we deduce that  $S_t w_0(x)$  is a  $C^1$  classical solution of (1) on  $]0, T] \times \mathbb{R}^N$ , continuous on  $[0, T] \times \mathbb{R}^N$ , and with initial data  $w_0$ . In turn, this implies that the function

$$w(t, x) := S_t w_0(x) + T \cdot H(0) \quad (t, x) \in [0, T] \times \mathbb{R}^N, \quad (106)$$

is also a  $C^1$  classical solution of (1) on  $]0, T] \times \mathbb{R}^N$ , continuous on  $[0, T] \times \mathbb{R}^N$ , and that satisfies

$$w(T, \cdot) \in \text{Lip}(\mathbb{R}^N), \quad \text{Lip}[w(T, \cdot)] \leq m, \quad (107)$$

$$w(T, x) = 0 \quad \forall x \in \mathbb{R}^N \setminus [-L, L]^N. \quad (108)$$

Next, notice that, by the above observations, the function

$$u(t, x) := -w(T - t, -x) \quad (109)$$

is a  $C^1$  classical solution of (1) on  $]0, T[ \times \mathbb{R}^N$ , continuous on  $[0, T] \times \mathbb{R}^N$ . Thus, recalling Remark 3, we deduce that  $u(t, x)$  is a viscosity solution of (1) on  $[0, T] \times \mathbb{R}^N$ , so that, by the uniqueness property (i) of the semigroup map  $S_t$ , one has

$$u(t, x) = S_t u_0, \quad u_0 := u(0, \cdot) \quad \forall (t, x) \in [0, T] \times \mathbb{R}^N. \quad (110)$$

Moreover, by virtue of (98), (107), (108), and by definition (3) it follows that

$$u_0 = -w(T, -\cdot) \in \mathcal{C}_{[L, M]}. \quad (111)$$

On the other hand, because of (102), (106), (109), there holds

$$S_T u_0(x) = -w_0(-x) - T \cdot H(0) = \psi(x) - T \cdot H(0) \quad \forall x \in \mathbb{R}^N. \quad (112)$$

Hence, (111)-(112) together yield

$$\psi \in S_T(\mathcal{C}_{[L, M]}) + T \cdot H(0), \quad (113)$$

which completes the proof of the proposition, being  $\psi$  an arbitrary element satisfying (101).  $\square$

**Remark 5.** The above proof shows that, for any given  $\psi \in \mathcal{SC}_{[L/2, m, K]}$ , one actually finds  $u_0 \in \mathcal{C}_{[L, m]}$  which is semiconvex with constant  $-\frac{1}{\alpha T}$  so that  $\psi = S_T u_0 + T \cdot H(0)$ . Indeed, by (106), (109), (110), one has  $-u_0 = S_T w_0(-\cdot) + T \cdot H(0)$ , which is semiconcave in  $\mathbb{R}^N$  with constant  $\frac{1}{\alpha T}$  thanks to Lemma 1.

*Proof of Lemma 2.* Observe first that, by Lemma 1, the map  $x \mapsto S_t u_0(x)$  is semiconcave for any fixed  $t \in ]0, T]$ . Therefore, once we establish the property (i) of Lemma 2, invoking Proposition 3 we immediately deduce that also the property (ii) holds. On the other hand, by the semiconcavity of  $S_t u_0$ , we know that  $S_t u_0$  is a continuous map. Hence, in order to prove the lemma, we only have to show that, for any fixed  $t \in ]0, T]$ , the map  $u(t, x) := S_t u_0(x)$  satisfies the lower bound

$$u(t, x + h) + u(t, x - h) - 2u(t, x) \geq -K_M \cdot |h|^2 \quad \forall x, h \in \mathbb{R}^N, \quad (114)$$

for some constant  $K_M > 0$ , depending on  $K$  and  $M$ .

1. Towards a proof of (114), fix  $x, h \in \mathbb{R}^N$ , and let  $y_h^\pm$  be a minimizer of the function

$$y \mapsto t \cdot H^*\left(\frac{x \pm h - y}{t}\right) + u_0(y) \quad (y \in \mathbb{R}^N), \quad (115)$$

where  $H^*$  denotes the Legendre transform of  $H$ . Then, recalling the Hopf-Lax formula (21), one has

$$u(t, x \pm h) = t \cdot H^*\left(\frac{x \pm h - y_h^\pm}{t}\right) + u_0(y_h^\pm). \quad (116)$$

Moreover, since  $y_h^\pm$  is a minimizer of (115), by the definition of the subdifferential in (12) it follows that there will be some

$$p_h^\pm \in D^- u_0(y_h^\pm), \quad (117)$$

such that

$$\nabla H^*\left(\frac{x \pm h - y_h^\pm}{t}\right) = p_h^\pm. \quad (118)$$

Since  $\text{Lip}[u_0] \leq M$ , applying Theorem 3-(iii) it follows that

$$|p_h^\pm| \leq M. \quad (119)$$

On the other hand, the Hopf-Lax formula implies that

$$u(t, x) \leq t \cdot H^*\left(\frac{x - \frac{y_h^+ + y_h^-}{2}}{t}\right) + u_0\left(\frac{y_h^+ + y_h^-}{2}\right). \quad (120)$$

Hence, combining (116), (120), we find

$$\begin{aligned} u(t, x + h) + u(t, x - h) - 2u(t, x) &\geq u_0(y_h^+) + u_0(y_h^-) - 2u_0\left(\frac{y_h^+ + y_h^-}{2}\right) + \\ &+ t \cdot \left[ H^*\left(\frac{x + h - y_h^+}{t}\right) + H^*\left(\frac{x - h - y_h^-}{t}\right) - 2 \cdot H^*\left(\frac{x - \frac{y_h^+ + y_h^-}{2}}{t}\right) \right]. \end{aligned} \quad (121)$$

Since  $H^*$  is convex and  $u_0$  is semiconvex with constant  $-K$ , we obtain from (121) the inequality

$$u(t, x + h) + u(t, x - h) - 2u(t, x) \geq -\frac{K}{4} \cdot |y_h^+ - y_h^-|^2. \quad (122)$$

**2.** In order to recover the estimate (114) from (122) we need to provide an upper bound on  $|y_h^+ - y_h^-|^2$ . To this end, observe first that, in view of (118), one has

$$\left\langle \nabla H^*\left(\frac{x+h-y_h^+}{t}\right) - \nabla H^*\left(\frac{x-h-y_h^-}{t}\right), -\frac{y_h^+ - y_h^-}{t} \right\rangle = \left\langle p_h^+ - p_h^-, -\frac{y_h^+ - y_h^-}{t} \right\rangle. \quad (123)$$

On the other hand, owing to the semiconvexity of  $u_0$  and by virtue of (14), (117), we can apply Proposition 2-(i) to get

$$\left\langle -p_h^+ + p_h^-, \frac{y_h^+ - y_h^-}{t} \right\rangle \leq \frac{K}{t} \cdot |y_h^+ - y_h^-|^2,$$

which, together with (123), yields

$$\left\langle \nabla H^*\left(\frac{x+h-y_h^+}{t}\right) - \nabla H^*\left(\frac{x-h-y_h^-}{t}\right), -\frac{y_h^+ - y_h^-}{t} \right\rangle \leq \frac{K}{t} \cdot |y_h^+ - y_h^-|^2. \quad (124)$$

Next, observe that there holds

$$\begin{aligned} \left\langle \nabla H^*\left(\frac{x+h-y_h^+}{t}\right) - \nabla H^*\left(\frac{x-h-y_h^-}{t}\right), -\frac{y_h^+ - y_h^-}{t} \right\rangle &= \\ &= \int_0^1 \left\langle D^2 H^*(z_\tau) \left( \frac{2h - (y_h^+ - y_h^-)}{t} \right), -\frac{y_h^+ - y_h^-}{t} \right\rangle d\tau, \end{aligned} \quad (125)$$

where

$$z_\tau = \tau \cdot \frac{x+h-y_h^+}{t} + (1-\tau) \cdot \frac{x-h-y_h^-}{t}. \quad (126)$$

Now, relying on (24), (118), (119), and assumption **(H2)**, we get

$$|z_\tau| \leq \sup_{|p| \leq M} |(\nabla H^*)^{-1}(p)| \leq \sup_{|p| \leq M} \|D^2 H(p)\| \cdot M = \alpha_M \cdot M \quad \forall \tau \in [0, 1],$$

where  $\alpha_M$  denotes a constant defined as in (100). Hence, one has

$$\int_0^1 \left\langle D^2 H^*(z_\tau) \left( \frac{2h}{t} \right), -\frac{y_h^+ - y_h^-}{t} \right\rangle d\tau \geq -2\beta_M \cdot \frac{|h| \cdot |y_h^+ - y_h^-|}{t^2}, \quad (127)$$

where

$$\beta_M := \sup_{|p| \leq \alpha_M \cdot M} \|D^2 H^*(p)\|. \quad (128)$$

On the other hand, notice that the definition of  $\alpha_M$  in (100) implies  $DH^2(p) \leq \alpha_M \cdot \mathbb{I}_N$  for  $|p| \leq M$ . Thus, recalling (24), we deduce that, for every  $q = (\nabla H^*)^{-1}(p)$ , with  $|p| \leq M$ , one has

$$D^2 H^*(q) = (D^2 H(p))^{-1} \geq \frac{1}{\alpha_M} \cdot \mathbb{I}_N,$$

which, by (118), (119), and because of the definition (126) of  $z_\tau$ , implies

$$D^2 H^*(z_\tau) \geq \frac{1}{\alpha_M} \cdot \mathbb{I}_N \quad \forall \tau \in [0, 1].$$



Therefore, we find

$$\int_0^1 \left\langle D^2 H^*(z_\tau) \left( -\frac{y_h^+ - y_h^-}{t} \right), -\frac{y_h^+ - y_h^-}{t} \right\rangle d\tau \geq \frac{1}{\alpha_M} \cdot \frac{|y_h^+ - y_h^-|^2}{t^2}. \quad (129)$$

Combining (125) with the lower bounds (127), (129), one obtains

$$\begin{aligned} & \left\langle \nabla H^* \left( \frac{x + h - y_h^+}{t} \right) - \nabla H^* \left( \frac{x - h - y_h^-}{t} \right), -\frac{y_h^+ - y_h^-}{t} \right\rangle \\ & \geq \frac{1}{\alpha_M} \cdot \frac{|y_h^+ - y_h^-|^2}{t^2} - 2\beta_M \cdot \frac{|h| \cdot |y_h^+ - y_h^-|}{t^2}. \end{aligned} \quad (130)$$

**3.** The upper bound (124) together with the lower bound (130) yields

$$2\beta_M \cdot |h| \cdot |y_h^+ - y_h^-| \geq \left( \frac{1}{\alpha_M} - Kt \right) |y_h^+ - y_h^-|^2.$$

In turn, recalling (98), from the above inequality it follows that

$$|y_h^+ - y_h^-| \leq 4\beta_M \alpha_M \cdot |h|.$$

Finally, using this last estimate, it is immediate to deduce (114) from (122), with  $K_M = 4K\beta_M^2\alpha_M^2$ , where  $\alpha_M, \beta_M$  are defined in (100) and (128), respectively. This completes the proof of the lemma.  $\square$

## 4.2 Part 2: Lower compactness estimates on a class of bump functions

We provide here a lower bound on the  $\varepsilon$ -entropy for the class of semiconcave functions  $\mathcal{SC}_{[L,M,K]}$  introduced in (51).

**Proposition 10.** *Given any  $L, M, K > 0$ , for every*

$$0 < \varepsilon \leq \min\{K, M\} \cdot \frac{\omega_N L^N}{(N+1)2^{N+8}}, \quad (131)$$

*there holds*

$$\mathcal{H}_\varepsilon(\mathcal{SC}_{[L,M,K]} \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) \geq \beta_{[L,K,N]}^{\mathcal{SC}} \cdot \frac{1}{\varepsilon^N}, \quad (132)$$

*where*

$$\beta_{[L,K,N]}^{\mathcal{SC}} := \frac{1}{8 \cdot \ln 2} \cdot \left( \frac{K \omega_N L^{N+1}}{48(N+1)2^{N+1}} \right)^N. \quad (133)$$

*Proof.* The proof is given in three steps. We shall first define a prototype  $C^1$  bump function with Lipschitz continuous gradient. Next, we shall consider a class of semiconcave functions  $\mathcal{U}_n$  defined as superpositions of such a bump function, localized on the  $N$ -dimensional cubes of a partition of the domain  $[-L, L]^N$ . Finally, we shall derive an optimal lower bound on the covering number  $\mathcal{N}_\varepsilon(\mathcal{U}_n \mid \mathbf{W}^{1,1}(\mathbb{R}^N))$  for a suitable choice of  $n$ , which then yields (132).

*Step 1: construction of a bump function.*

Consider the continuously differentiable function  $c : [0, 1] \rightarrow \mathbb{R}$  defined by

$$c(r) = \begin{cases} \left(\frac{1}{4}\right)^2 - \int_0^r \left(\frac{1}{4} - |\frac{1}{4} - s|\right) ds & \text{if } r \in [0, \frac{1}{2}], \\ 0 & \text{if } r \in [\frac{1}{2}, 1]. \end{cases} \quad (134)$$

Then, we compute

$$c'(r) = \begin{cases} |\frac{1}{4} - r| - \frac{1}{4} & \text{if } r \in [0, \frac{1}{2}], \\ 0 & \text{if } r \in [\frac{1}{2}, 1]. \end{cases}$$

Thus,  $c'$  is Lipschitz continuous with Lipschitz constant 1 and there holds

$$|c'(r)| \leq r \quad \forall r \in [0, 1]. \quad (135)$$

Moreover, one has

$$\|c\|_{\mathbf{L}^\infty([0,1])} \leq \frac{1}{16}, \quad \|c'\|_{\mathbf{L}^\infty([0,1])} \leq \frac{1}{4}. \quad (136)$$

We now proceed to construct our bump function  $b : [-L, L]^N \rightarrow \mathbb{R}$  as follows:

$$b(x) = \begin{cases} \frac{KL^2}{6} c\left(\frac{|x|}{L}\right) & \text{if } x \in B(0, \frac{L}{2}) \\ 0 & \text{if } x \in [-L, L]^N \setminus B(0, \frac{L}{2}). \end{cases} \quad (137)$$

One can check that

$$\nabla b(x) = \frac{KL}{6} c'\left(\frac{|x|}{L}\right) \frac{x}{|x|} \quad \forall x \in [-L, L]^N \setminus 0 \quad \text{and} \quad \nabla b(0) = 0. \quad (138)$$

Thus, because of (136), there holds

$$\|b\|_{\mathbf{L}^\infty([-L, L]^N)} \leq \frac{KL^2}{96} \quad \|\nabla b\|_{\mathbf{L}^\infty([-L, L]^N)} \leq \frac{KL}{24}. \quad (139)$$

Furthermore, since  $c'$  is 1-Lipschitz, observing that

$$\left| \frac{y}{|y|} - \frac{x}{|x|} \right| \leq \frac{2|y - x|}{|x|} \quad \forall x, y \neq 0,$$

and relying on (135), (138), it follows that  $\nabla b$  is Lipschitz continuous with constant  $K/2$  in  $[-L, L]^N$ . On the other hand, observing that

$$\nabla b(x) = 0 \quad \forall x \in [-L, L]^N \setminus B(0, \frac{L}{2}), \quad (140)$$

a straightforward computation shows that

$$\begin{aligned} \|\nabla b\|_{\mathbf{L}^1([-L, L]^N)} &= \frac{KL}{6} \int_{B(0, L/2)} \left| c'\left(\frac{|x|}{L}\right) \right| dx \\ &= \frac{KL^{N+1}}{6} \int_{B(0, 1/2)} |c'(|x|)| dx = \frac{NKL^{N+1}\omega_N}{6} \int_0^1 |c'(r)| r^{N-1} dr \\ &= \frac{KL^{N+1}\omega_N}{6} \cdot \frac{2^N - 1}{2(N+1)4^N}. \end{aligned}$$

Thus, setting

$$\beta_{[L, K, N]} := \frac{KL^{N+1}\omega_N}{12} \cdot \frac{2^N - 1}{(N+1)4^N}, \quad (141)$$

we have

$$\|\nabla b\|_{\mathbf{L}^1([-L, L]^N)} = \beta_{[L, K, N]}. \quad (142)$$

Now, given any positive integer  $n \in \mathbb{N}$ , let us consider the continuously differentiable function  $b_n : [-\frac{L}{n}, \frac{L}{n}]^N \rightarrow \mathbb{R}$  defined as

$$b_n(x) = \frac{b(nx)}{n^2}, \quad \forall x \in \left[-\frac{L}{n}, \frac{L}{n}\right]^N. \quad (143)$$

Thus, by (137) one has

$$b_n(x) = 0 \quad \text{if} \quad x \in \left[-\frac{L}{n}, \frac{L}{n}\right]^N \setminus B(0, \frac{L}{2n}). \quad (144)$$

Noting that  $\nabla b_n(x) = \frac{1}{n} \cdot \nabla b(nx)$  for  $x \in [-\frac{L}{n}, \frac{L}{n}]^N$ , and relying on (139), (142), one can easily check that

$$\|\nabla b_n\|_{\mathbf{L}^\infty([- \frac{L}{n}, \frac{L}{n}]^N)} \leq \frac{KL}{24n}, \quad \|\nabla b_n\|_{\mathbf{L}^1([- \frac{L}{n}, \frac{L}{n}]^N)} = \frac{1}{n^{N+1}} \|\nabla b\|_{\mathbf{L}^1([-L, L]^N)} = \frac{\beta_{[L, K, N]}}{n^{N+1}}. \quad (145)$$

Moreover, since  $\nabla b$  is Lipschitz continuous with constant  $K/2$ , we have that  $\nabla b_n$  is also Lipschitz continuous with constant  $K/2$ . By Remark 1 this implies that  $b_n$  and  $-b_n$  are semiconcave functions with constant  $K$ .

*Step 2: a class of semiconcave functions defined as superpositions of bump functions.*

For any integer  $n \geq 1$  let us divide  $[-L, L]^N$  into  $n^N$  cubes of side  $\frac{2L}{n}$  as in the proof of Proposition 7. More precisely, we shall use the notation

$$[-L, L]^N = \bigcup_{\iota \in \{1, \dots, n\}^N} \square_\iota, \quad (146)$$

where  $\iota = (\iota_1, \dots, \iota_N) \in \{1, \dots, n\}^N$  is a multiindex and

$$\square_\iota := (-L, \dots, -L) + \frac{L}{n} \iota + \left[-\frac{L}{n}, \frac{L}{n}\right]^N$$

is an  $N$ -dimensional cube centered at  $x_\iota := (-L, \dots, -L) + \frac{L}{n} \iota$ . Let us now adapt our bump function  $b_n$  in (143) to the cube  $\square_\iota$  defining

$$b_n^\iota(x) = \begin{cases} b_n(x - x_\iota) & \text{if } x \in \square_\iota, \\ 0 & \text{if } x \in \mathbb{R}^N \setminus \square_\iota. \end{cases}$$

One can easily verify that the continuously differentiable function  $b_n^\iota : \mathbb{R}^N \rightarrow \mathbb{R}$  shares the same properties of  $b_n$ . In particular, by (144), (145), there holds:

- (i)  $b_n^\iota(x) = 0$  for all  $x \in \mathbb{R}^N \setminus B(x_\iota, \frac{L}{2n})$ ,
- (ii)  $\|\nabla b_n^\iota\|_{\mathbf{L}^\infty(\mathbb{R}^N)} \leq \frac{KL}{24n}$  and  $\|\nabla b_n^\iota\|_{\mathbf{L}^1(\mathbb{R}^N)} = \frac{\beta_{[L, K, N]}}{n^{N+1}}$ ,
- (iii)  $b_n^\iota$  and  $-b_n^\iota$  are semiconcave with constant  $K$ .

Next, we proceed to construct a class of semiconcave functions in the set  $\mathcal{SC}_{[L, M, K]}$ , defined as combinations of the bump functions  $b_n^\iota$ . Namely, consider the set of  $n^N$ -tuples

$$\Delta_n = \left\{ \delta = (\delta_\iota)_{\iota \in \{1, \dots, n\}^N} \mid \delta_\iota \in \{-1, 1\} \right\},$$

and, for every  $\delta = (\delta_\iota)_{\iota \in \{1, \dots, n\}^N} \in \Delta_n$ , define the function  $u_\delta : \mathbb{R}^N \rightarrow \mathbb{R}$  by setting

$$u_\delta := \sum_{\iota \in \{1, \dots, n\}^N} \delta_\iota b_n^\iota. \quad (147)$$

Observe that, by properties (i)-(iii) above, every function  $u_\delta$  has support contained in  $[-L, L]^N$ , is semiconcave with semiconcavity constant  $K$ , and satisfies  $\|\nabla u\|_{\mathbf{L}^\infty(\mathbb{R}^N)} \leq M$  provided that

$$n \geq \frac{KL}{24M}. \quad (148)$$

Therefore, recalling definition (51), one has

$$\mathcal{U}_n := \{u_\delta \mid \delta \in \Delta_n\} \subset \mathcal{SC}_{[K, L, M]}, \quad (149)$$

for all such  $n$ . Hence, in order to establish (132), it will be sufficient to show that there holds

$$\mathcal{H}_\varepsilon(\mathcal{U}_n \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) \geq \beta_{[L, K, N]}^{\mathcal{SC}} \cdot \frac{1}{\varepsilon^N} \quad (150)$$

for every  $\varepsilon$  sufficiently small and for a suitable choice of  $n$  satisfying (148).

*Step 3: estimate of the  $\varepsilon$ -entropy for superpositions of bump functions by a combinatorial argument.*

Towards an estimate of the covering number  $\mathcal{N}_\varepsilon(\mathcal{U}_n \mid \mathbf{W}^{1,1}(\mathbb{R}^N))$ , fix  $\bar{\delta} \in \Delta_n$ , and let us define the set of  $n^N$ -tuples

$$\mathcal{I}_{\bar{\delta}, n}(\varepsilon) = \left\{ \delta \in \Delta_n \mid \|\nabla u_{\bar{\delta}} - \nabla u_\delta\|_{\mathbf{L}^1(\mathbb{R}^N)} \leq \varepsilon \right\}.$$

Notice that, by construction, the cardinality of the set  $\mathcal{I}_{\bar{\delta}, n}(\varepsilon)$  is independent of the choice of  $\bar{\delta} \in \Delta_n$ . Let us denote it by

$$C_n(\varepsilon) := \#(\mathcal{I}_{\bar{\delta}, n}(\varepsilon)).$$

Moreover, any element of an  $\varepsilon$ -cover in  $\mathbf{W}^{1,1}$  of  $\mathcal{U}_n$  contains at most  $C_n(2\varepsilon)$  functions of  $\mathcal{U}_n$ . Hence, since the cardinality of  $\mathcal{U}_n$  is the same as the cardinality of  $\Delta_n$ , which is  $\#(\Delta_n) = 2^{n^N}$ , it follows that the number of sets in an  $\varepsilon$ -cover in  $\mathbf{W}^{1,1}$  of  $\mathcal{U}_n$  is at least

$$\mathcal{N}_\varepsilon(\mathcal{U}_n \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) \geq \frac{2^{n^N}}{C_n(2\varepsilon)}. \quad (151)$$

Aiming at an upper bound on  $C_n(2\varepsilon)$ , observe that for any given pair  $\delta, \bar{\delta} \in \Delta_n$ , one has

$$\|\nabla u_{\bar{\delta}} - \nabla u_\delta\|_{\mathbf{L}^1(\mathbb{R}^N)} = d(\bar{\delta}, \delta) \cdot 2\|\nabla b_n\|_{\mathbf{L}^1([-L/n, L/n]^N)}, \quad (152)$$

where

$$d(\bar{\delta}, \delta) := \#\left(\{\iota \in \{1, \dots, n\}^N \mid \bar{\delta}_\iota \neq \delta_\iota\}\right). \quad (153)$$

Thus, relying on (145), (152), we deduce that

$$d(\bar{\delta}, \delta) \leq \frac{n^{N+1}}{\beta_{[L, K, N]}} \cdot \varepsilon \quad \Longleftrightarrow \quad \|\nabla u_{\bar{\delta}} - \nabla u_\delta\|_{\mathbf{L}^1(\mathbb{R}^N)} \leq 2\varepsilon. \quad (154)$$

Hence, performing a standard combinatorial computation of the number of  $n^N$ -tuples that differ for a given number of entries, we find

$$C_n(2\varepsilon) \leq \sum_{l=0}^{\left\lfloor \frac{n^{N+1}}{\beta_{[L,K,N]}} \cdot \varepsilon \right\rfloor} \binom{n^N}{l}. \quad (155)$$

Next, observe that if  $X_1, \dots, X_{n^N}$  are independent random variables with uniform Bernoulli distribution  $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = \frac{1}{2}$ , then, for any  $k \leq n^N$ , one has

$$\sum_{l=0}^k \binom{n^N}{l} = 2^{n^N} \cdot \mathbb{P}(X_1 + \dots + X_{n^N} \leq k). \quad (156)$$

Now, set  $S_{n^N} = X_1 + \dots + X_{n^N}$ , and recall Hoeffding's inequality ([14, Theorem 2]) which guarantees that, for any  $\mu > 0$ ,

$$\mathbb{P}(S_{n^N} - \mathbb{E}[S_{n^N}] \leq -\mu) \leq \exp\left(-\frac{2\mu^2}{n^N}\right), \quad (157)$$

where  $\mathbb{E}[S_{n^N}]$  denotes the expectation of  $S_{n^N}$ . Since  $\mathbb{E}[S_{n^N}] = \frac{n^N}{2}$ , taking  $\mu = \frac{n^N}{2} - \left\lfloor \frac{n^{N+1}}{\beta_{[L,K,N]}} \cdot \varepsilon \right\rfloor$  and assuming

$$n \leq \frac{\beta_{[L,K,N]}}{2\varepsilon}, \quad (158)$$

from (155), (156) and (157) it follows that

$$\begin{aligned} C_n(2\varepsilon) &\leq 2^{n^N} \exp\left(-\frac{2\left(\frac{n^N}{2} - \left\lfloor \frac{n^{N+1}}{\beta_{[L,K,N]}} \cdot \varepsilon \right\rfloor\right)^2}{n^N}\right) \\ &\leq 2^{n^N} \exp\left(-\frac{n^N}{2} \left(1 - \frac{2n\varepsilon}{\beta_{[L,K,N]}}\right)^2\right). \end{aligned} \quad (159)$$

In turn, (159) together with (151), yields

$$\mathcal{N}_\varepsilon(\mathcal{U}_n \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) \geq \exp\left(\frac{n^N}{2} \left(1 - \frac{2n\varepsilon}{\beta_{[L,K,N]}}\right)^2\right) \quad (160)$$

for all  $n$  satisfying (158). Now, if we take

$$0 < \varepsilon \leq \min\left\{\frac{\beta_{[L,K,N]}}{8}, \frac{6M\beta_{[L,K,N]}}{KL}\right\}, \quad (161)$$

choosing

$$n_\varepsilon := \left\lfloor \frac{\beta_{[L,K,N]}}{4\varepsilon} \right\rfloor + 1, \quad (162)$$

one easily check that  $n_\varepsilon$  satisfies both bounds (148), (158). Hence, relying on (149), (160), we find the lower bound

$$\begin{aligned} \mathcal{N}_\varepsilon(\mathcal{SC}_{[K,L,M]} \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) &\geq \mathcal{N}_\varepsilon(\mathcal{U}_{n_\varepsilon} \mid \mathbf{W}^{1,1}(\mathbb{R}^N)) \\ &\geq \exp\left(\frac{n_\varepsilon^N}{8}\right) \geq \exp\left(\frac{\beta_{[L,K,N]}^N}{8(4\varepsilon)^N}\right), \end{aligned} \quad (163)$$

for all  $\varepsilon$  satisfying (161). In turn, this estimate yields (132) for all  $\varepsilon$  satisfying (131), taking  $\log_2$  of both sides of (163) and observing that, by (133), (141), one has

$$\beta_{[L,K,N]} \geq \frac{K \omega_N L^{N+1}}{24(N+1) 2^N}, \quad \frac{1}{8 \ln 2} \cdot \left( \frac{\beta_{[L,K,N]}}{4} \right)^N \geq \beta_{[L,K,N]}^{SC}.$$

□

### 4.3 Conclusion of the proof of Theorem 1-(ii)

Given,  $L, M, T > 0$ , combining Proposition 9 and Proposition 10 we find that, for every

$$0 < \varepsilon \leq \min \left\{ M, \frac{1}{4 \|D^2 H(0)\| \cdot T}, \frac{L}{4 \|D^2 H(0)\| \cdot T} \right\} \cdot \frac{\omega_N L^N}{(N+1) 2^{N+8}}, \quad (164)$$

there holds

$$\mathcal{H}_\epsilon \left( S_T(\mathcal{C}_{[L,M]}) + T \cdot H(0) \mid \mathbf{W}^{1,1}(\mathbb{R}^N) \right) \geq \beta_{\left[ \frac{L}{2}, \frac{1}{4 \|D^2 H(0)\| \cdot T}, N \right]}^{SC} \cdot \frac{1}{\varepsilon^N}, \quad (165)$$

where

$$\beta_{\left[ \frac{L}{2}, \frac{1}{4 \|D^2 H(0)\| \cdot T}, N \right]}^{SC} = \frac{1}{8 \cdot \ln 2} \cdot \left( \frac{\omega_N L^{N+1}}{3(N+1) 2^{(2N+8)} \cdot \|D^2 H(0)\| \cdot T} \right)^N. \quad (166)$$

This establishes the lower bound (7). □

## Acknowledgements

This work was partially supported by the National Group for Mathematical Analysis and Probability (GNAMPA) of the Istituto Nazionale di Alta Matematica “Francesco Severi” (INdAM), the CNRS-INdAM European Research Group (GDRE) on Control of Partial Differential Equations (CONEDP), and the European Union 7th Framework Programme [FP7-PEOPLE-2010-ITN] under grant agreement n.264735-SADCO. Fabio Ancona was partially supported by the Miur-Prin 2012 Project “Nonlinear Hyperbolic Partial Differential Equations, Dispersive and Transport Equations: theoretical and applicative aspects” and by Fondazione CaRiPaRo Project “Nonlinear Partial Differential Equations: models, analysis, and control-theoretic problems”.

## References

- [1] G. Acosta and R. C. Dúran, An optimal Poincaré inequality in  $L^1$  for convex domains. Proc. Amer. Math. Soc. Vol 132 (2003), no.1, 195-202.
- [2] G. Alberti and L. Ambrosio, A geometrical approach to monotone functions in  $\mathbb{R}^N$ . Math. Z. 230 (1999), 259-316.
- [3] L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, Oxford Science Publications, Clarendon Press, Oxford, UK, (2000).
- [4] F. Ancona, O. Glass and K. T. Nguyen, Lower compactness estimates for scalar balance laws, Comm. Pure Appl. Math 65 (2012), no. 9, 1303-1329.

- [5] F. Ancona, O. Glass and K. T. Nguyen, On compactness estimates for hyperbolic systems of conservation laws, preprint (2014).
- [6] F. Ancona, O. Glass and K. T. Nguyen, On quantitative compactness estimates for hyperbolic conservation laws, to appear on: Hyperbolic problems: theory, numerics and applications. Proceedings of the 14th International Conference on Hyperbolic Problems (HYP2012), AIMS, Springfield, MO, 2014.
- [7] M. Bardi and I. Capuzzo Dolcetta, Optimal control and viscosity solutions of Hamilton-Jacobi equations. Birkhäuser, Boston (1997).
- [8] P. L. Bartlett, S. R. Kulkarni and S.E. Posner, Covering numbers for real-valued function classes. IEEE Trans. Inform. Theory 43 (1997), no. 5, 1721-1724.
- [9] P. Cannarsa and C. Sinestrari, Semiconcave functions, Hamilton-Jacobi equations, and optimal control. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston, 2004.
- [10] M.G. Crandall and P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations, Trans. Amer. Math. Soc. 277 (1983), no. 1, 1-42.
- [11] M.G. Crandall, L.C. Evans and P.-L. Lions, Some properties of viscosity solutions of Hamilton-Jacobi equations, Trans. Amer. Math. Soc. 282 (1984), 487-502.
- [12] C. De Lellis and F. Golse, A Quantitative Compactness Estimate for Scalar Conservation Laws, Comm. Pure Appl. Math. 58 (2005), no. 7, 989-998.
- [13] L.C. Evans, Partial Differential Equations, Graduate Studies in Mathematics, vol. 319, AMS, Providence, 1991.
- [14] W. Hoeffding, Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58 (1963), 13-30.
- [15] K.H. Karlsen and N.H. Risebro, A note on front tracking and equivalence between viscosity solutions of Hamilton-Jacobi equations and entropy solutions of scalar conservation laws. Nonlinear Anal. 50 (2002), no. 4, Ser. A: Theory Methods, 455-469.
- [16] A.N. Kolmogorov and V.M. Tikhomirov,  $\varepsilon$ -Entropy and  $\varepsilon$ -capacity of sets in functional spaces. Uspekhi Mat. Nauk 14 (1959), 3-86.
- [17] P. D. Lax, Accuracy and resolution in the computation of solutions of linear and nonlinear equations. Recent advances in numerical analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1978). Publ. Math. Res. Center Univ. Wisconsin, 107-117. Academic Press, New York, 1978.
- [18] P.D. Lax, Course on hyperbolic systems of conservation laws.. XXVII Scuola Estiva di Fis. Mat., Ravello, 2002.
- [19] O. A. Oleinik, Discontinuous solutions of non-linear differential equations. Uspekhi Mat. Nauk (N.S.) 12 (1957) no. 3(75), 3-73. (Russian) English translation in Ann. Math. Soc. Trans. Ser. 2 26, 95-172.
- [20] B. Temple, No  $L^1$ -contractive metrics for systems of conservation laws. Trans. Amer. Math. Soc. 288 (1985), 471-480.